

Názov prednášky:

Jadrové analýzy hustoty, Geograficky vážená regresia

Osnova prednášky:

Jadrové analýzy hustoty

- „population parameter“
- zohľadnenie váhy

Geograficky vážená regresia

Odporúčaná literatúra

KAŇUK, J., 2015: Priestorové analýzy a modelovanie. Vysokoškolské učebné texty. Prírodovedecká fakulta UPJŠ v Košiciach, 106 s.

FOTHERINGHAM, A. M., CHARLTON, M. E. AND BRUNSDON, C. (1998). Geographically weighted regression: a natural evolution of the expansion method for spatial data analysis. Environment and Planning A, Vol. 30, p. 1905-1927.

LLOYD, C.D. 2009: Local methods for spatial analysis, CRC Press.

YEE LEUNG, CHANG-LIN MEI, WEN-XIU ZHANG (2000): Statistical tests for spatial nonstationarity based on the geographically weighted regression model. Environment and Planning A, Vol. 32, p. 9-32

Jadrové analýzy hustoty – teoretický úvod

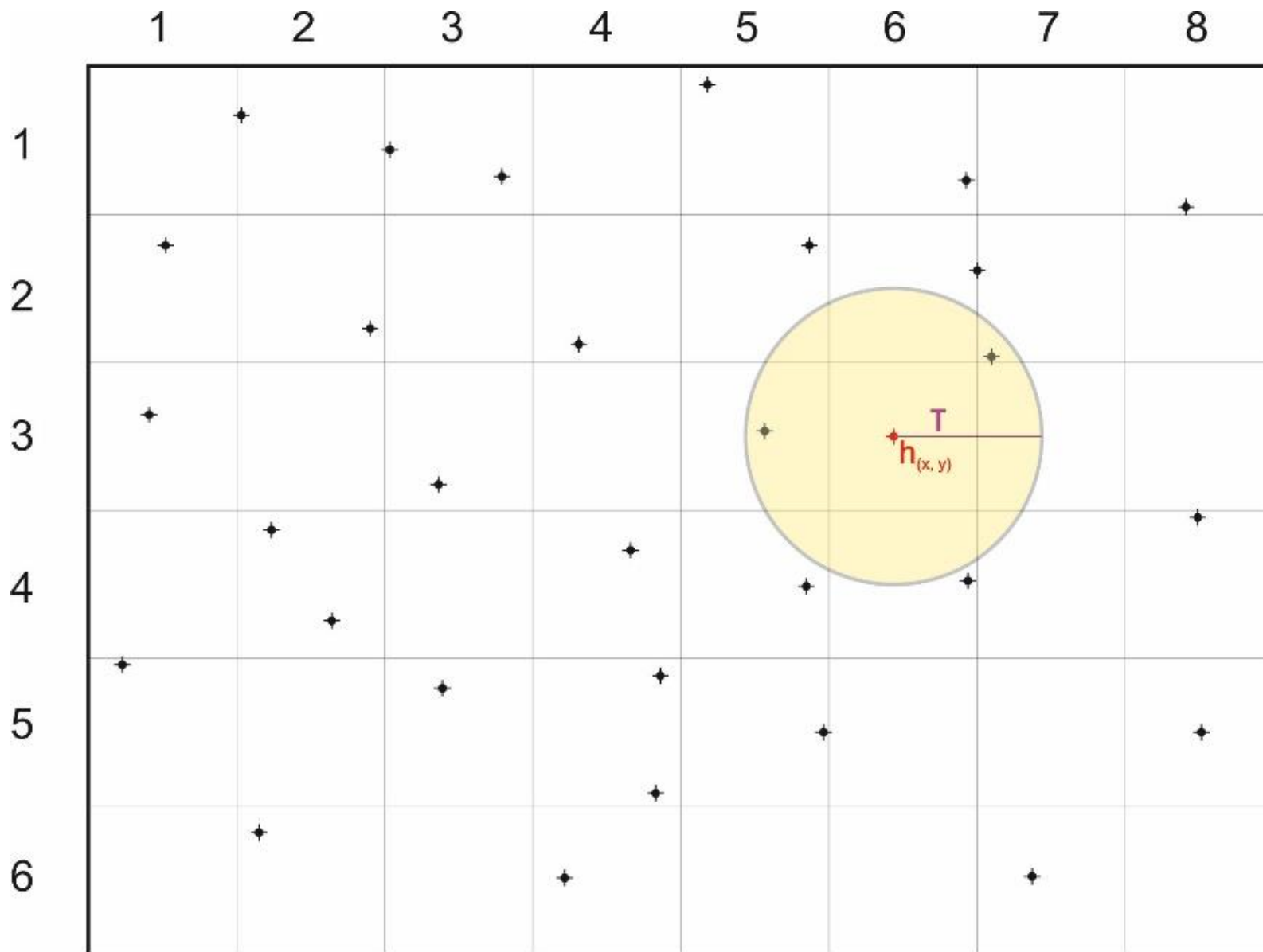
- AJ – Kernel density estimation
- je pomerne sofistikovaná metóda, pomocou ktorej je možné skúmať priestorové rozdiely v hustote výskytu javu
- hustotu výskytu javu by sme mohli definovať ako podiel počtu výskytov javu na jednotku plochy
- v prípade obyvateľstva sa často používa charakteristika hustoty zaľudnenia, ktorá vyjadruje počet obyvateľov na rozlohu administratívnej jednotky (štátu, kraja, okresu, a pod.). Tento prístup sa však nedá aplikovať na všetky geopriestorové javy (napr. kategorizáciu územia na základe morfometrickej charakteristiky relatívneho výškového prevýšenie územia, koncentrácia (hustota) vybraného odvetia hospodárstva, hodnotenie štruktúry obyvateľstva a pod.)
- je potrebné stanoviť si určité územie, pre ktoré budeme počítat hustotu javu → „**sledovaná oblasť**“ (*kernel* – jadro)
- analýza hustoty je založená na hodnotení hustoty pre malé regióny (kernely).
- je potrebné definovať **tvar** a **veľkosť** sledovanej oblasti
- **tvar** sledovanej oblasti môže byť rôzny. Najčastejšie sa používa kruh, obdĺžnik, avšak je možné použiť aj komplikovanejšie geometrické tvary, napr. rôzno-tvaré polygóny, medzikružie, trojuholníky, atď.
- dokonca je možné uvažovať o rôznych tvaroch pre jeden dataset, v ktorom sú záznamy o výskyte sledovaného javu.
- **veľkosť** sledovanej oblasti sa definuje spravidla pomocou **polomeru vyhľadávania (*bandwidth*)**.
- veľkosť sledovanej oblasti môže byť taktiež variabilná.
- Výsledné hodnoty hustoty sledovaného javu sú spravidla vyjadrené pomocou **rastrovej dátovej reprezentácie**.

Jadrové analýzy hustoty – princíp

Vstupné bodové pole (čierne body) s rastrom (štvorčeky) s veľkosťou 1m x1 m.

Hustota bodov $h_{(x,y)}$ sa počíta pre stredy každej bunky rastra.

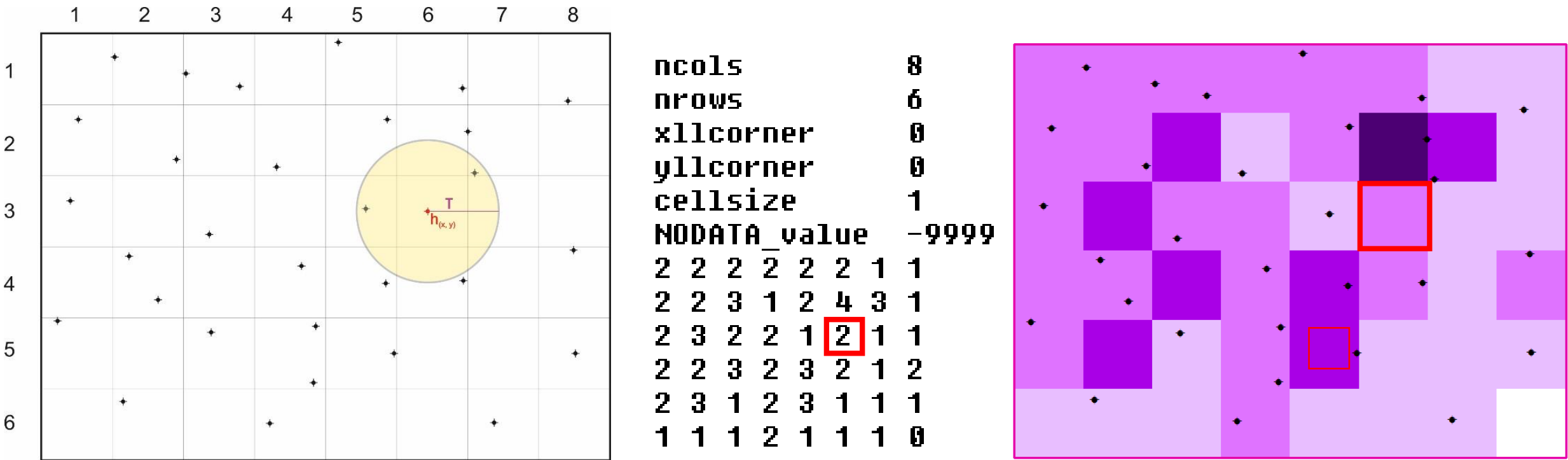
V našom prípade do sledovanej plochy zasahujú 2 body.



Za sledovanú oblasť bol zvolený kruh s polomerom τ 1 m, čiže veľkosť plochy kruhu je $S_{(x,y)} = 3,14157 \text{ m}^2$.

Jadrové analýzy hustoty – krok 1

Prvým krokom je určenie počtu bodov v sledovanej oblasti



V našom prípade teda najskôr určíme počet bodov spadajúcich do kruhu so stredom kruhu v strede bunky rastra. Uvedený postup sa zopakuje pre všetky bunky rastra a výsledok sa zapíše do matice hodnôt.

Jadrové analýzy hustoty – krok 2

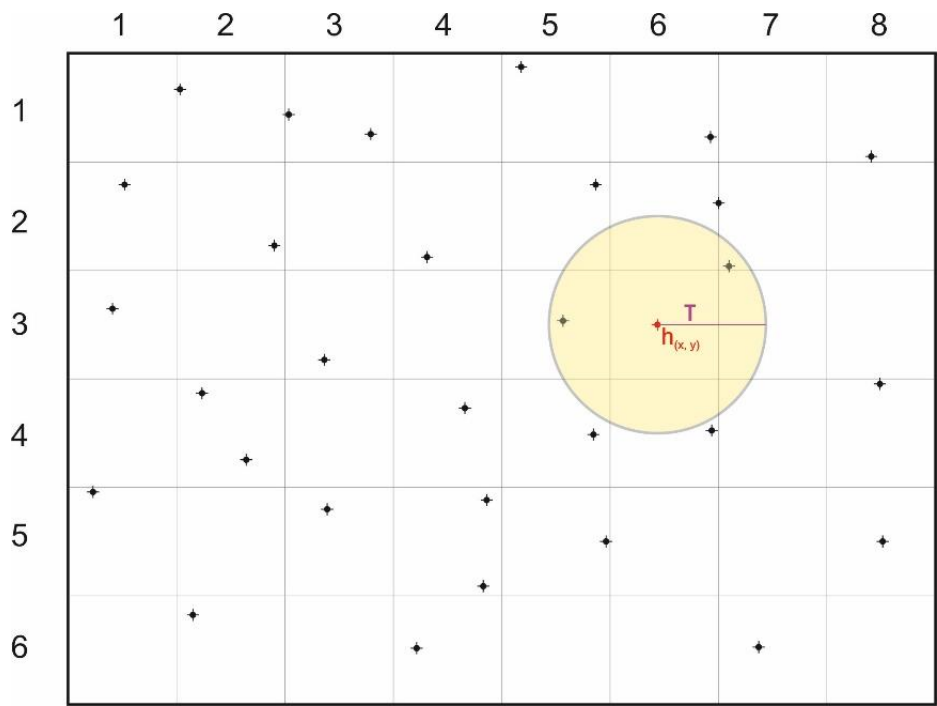
výpočet hustoty bodov pre sledovanú oblasť

$$h_{(x,y)} = \frac{n_{S(x,y)}}{S_{(x,y)}}$$

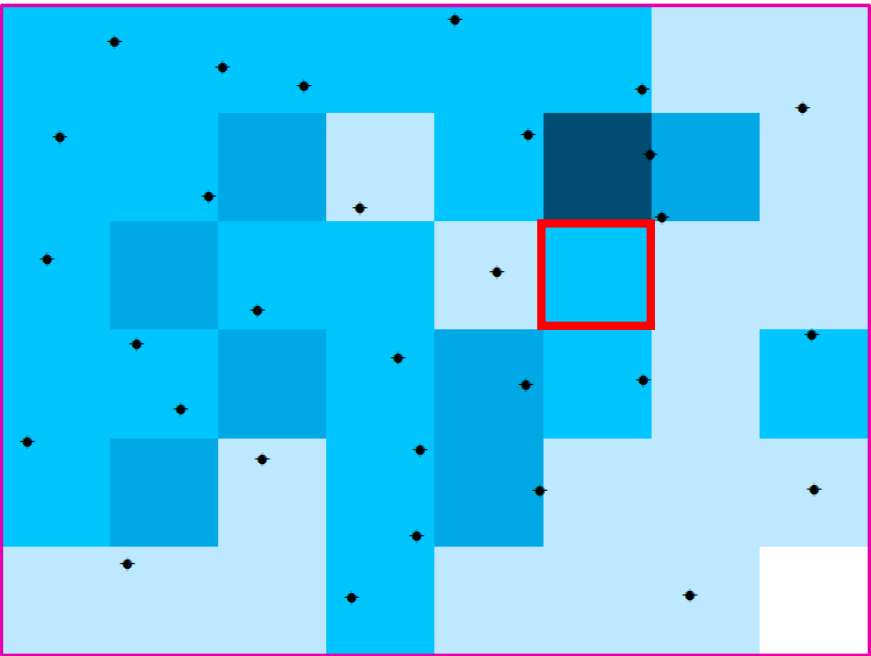
$h_{(x,y)}$ – hustota javu so súradnicami x a y

$n_{S(x,y)}$ - počet bodov v sledovanej ploche

$S_{(x,y)}$ - rozloha sledovanej oblasti

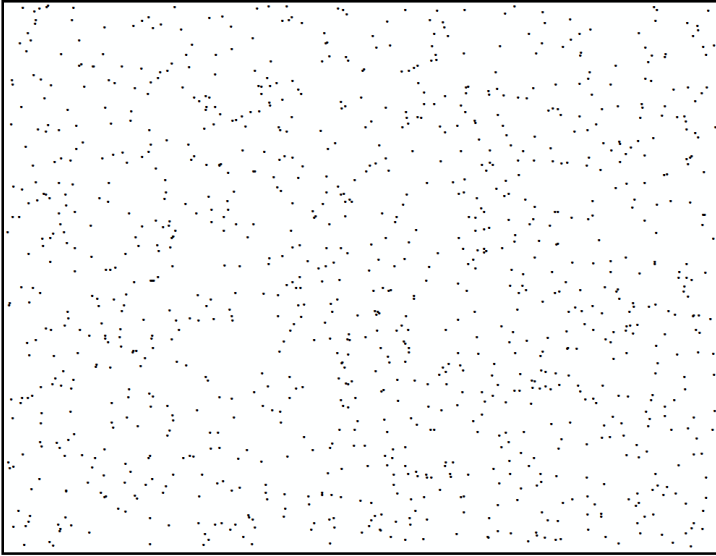


ncols	8							
nrows	6							
xllcorner	0							
yllcorner	0							
cellsize	1							
NODATA_value	-9999							
0,64	0,64	0,64	0,64	0,64	0,64	0,32	0,32	
0,64	0,64	0,95	0,32	0,64	1,27	0,95	0,32	
0,64	0,95	0,64	0,64	0,32	0,64	0,32	0,32	
0,64	0,64	0,95	0,64	0,95	0,64	0,32	0,64	
0,64	0,95	0,32	0,64	0,95	0,32	0,32	0,32	
0,32	0,32	0,32	0,64	0,32	0,32	0,32	0	

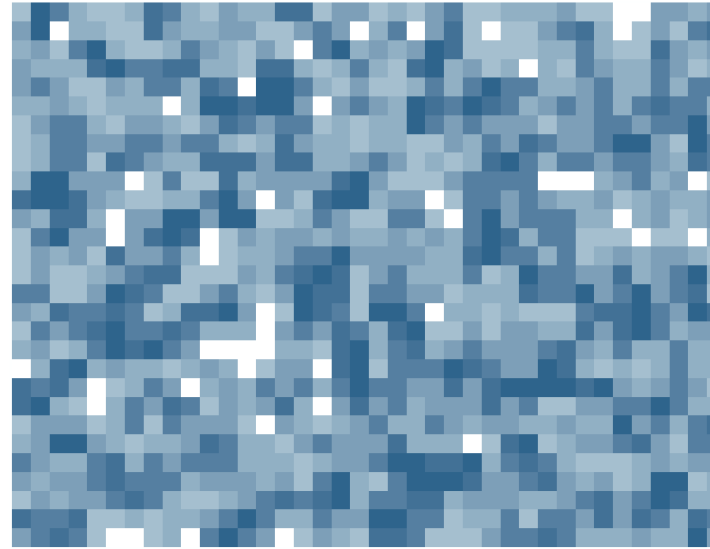


$$S_{(x,y)} = 3,14157 \text{ m}^2$$

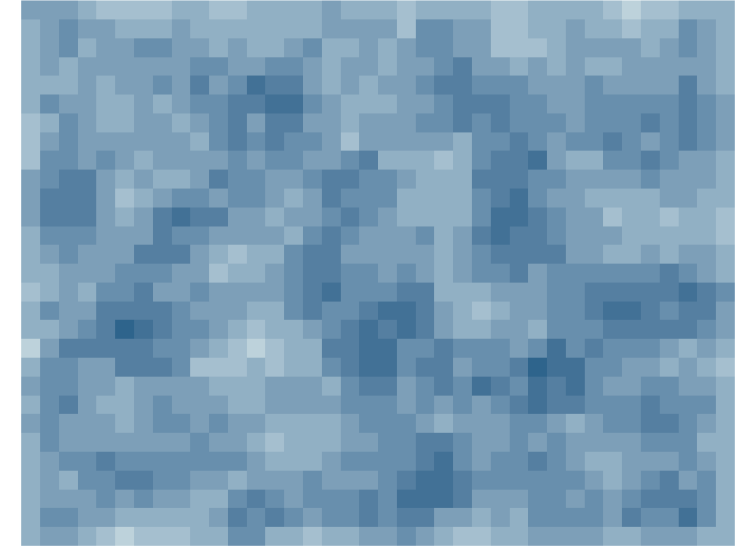
Jadrové analýzy hustoty – veľkosť sledovanej oblasti



Vstupné bodové pole



Polomer kruhu výhľadávania 50 m

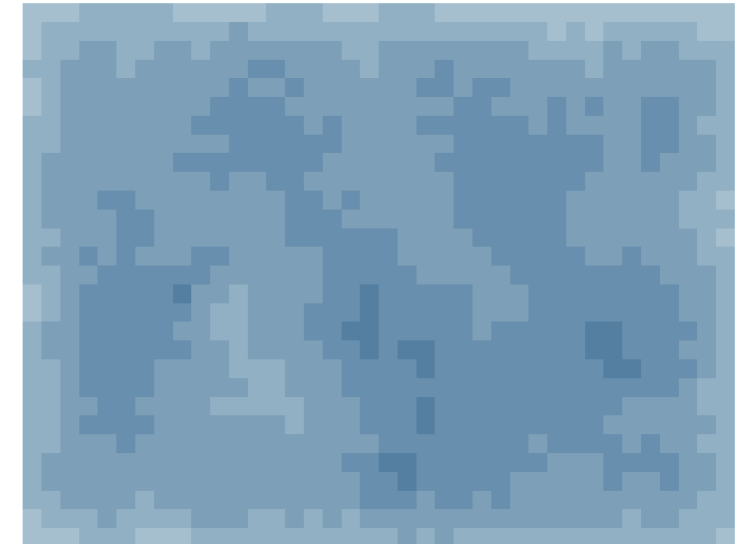
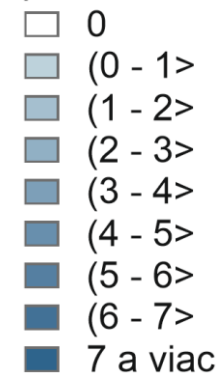


Polomer kruhu výhľadávania 100 m

Čím bude plocha sledovanej oblasti vyhľadávania rozľahlejšia (polomer kruhu bude väčší), tým je výsledný povrch hladší.

Naopak, čím je plocha sledovanej oblasti menšia (polomer kruhu menší), tým sú viac viditeľné zhluky (klastre) výskytov sledovaného javu.

počet bodov na 1 ha



Polomer kruhu výhľadávania 200 m

- je možné vypočítať hustotu javu na jednotku plochy, avšak nevýhodou tejto metódy je, že nezohľadňuje intenzitu (veľkosť) sledovaného javu, iba počet jeho výskytov.

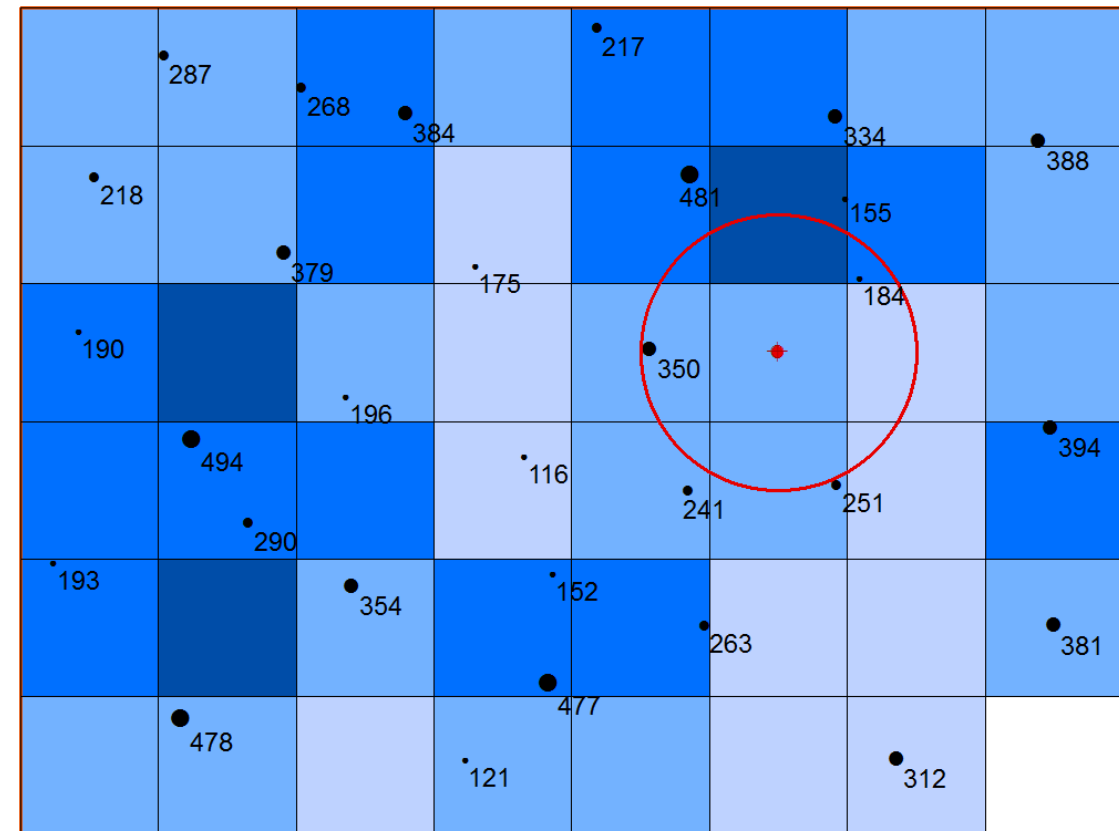
Jadrové analýzy hustoty – „population parameter“

Aby bolo možné zohľadniť intenzitu sledovaného javu, je zavedený parameter, ktorý sa vo väčšine GIS-ov označuje ako „population parameter“ p_i .

$$h_{(x,y)} = \frac{\sum_{i=1}^n p_i}{S_{(x,y)}}$$

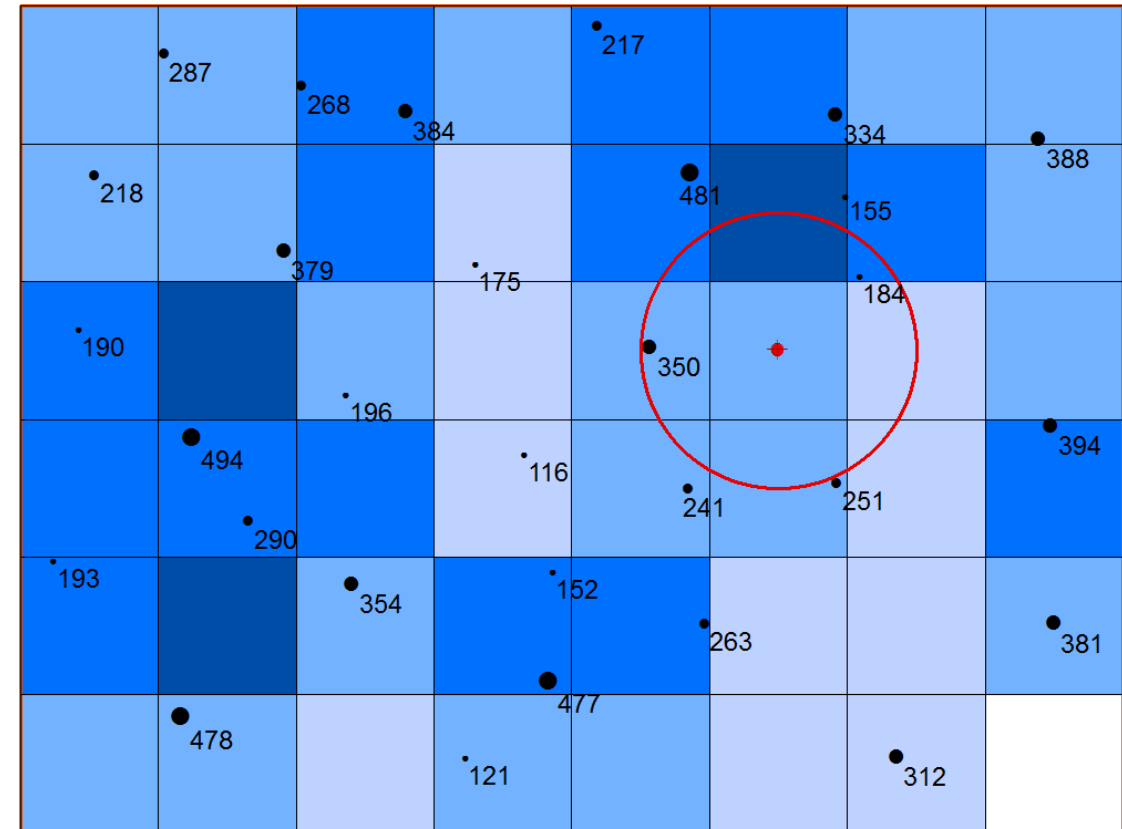
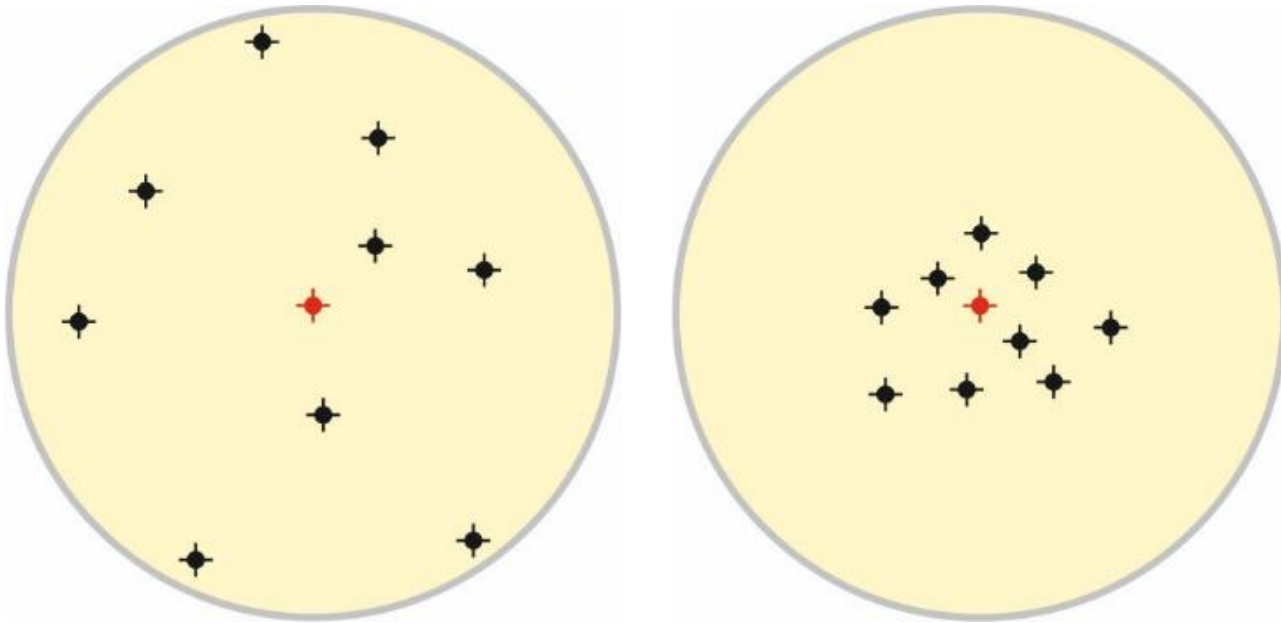
p_i – je „population parameter“

```
ncols      8
nrows      6
xllcorner  0
yllcorner  0
cellsize   1
NODATA_value -9999
160,75 176,66 207,54 191,30 222,18 259,42 106,32 123,50
129,87 190,03 298,57 55,70 264,52 367,33 214,22 123,50
217,72 340,27 183,03 92,63 111,41 169,98 58,57 125,41
218,68 249,55 267,38 85,31 162,02 156,61 79,90 246,69
213,59 357,14 112,68 200,22 283,93 83,72 99,31 121,28
152,15 152,15 38,52 190,35 151,83 99,31 99,31 0,00
```



Jadrové analýzy hustoty – „population parameter“

Aj táto metóda má isté obmedzenia, ktoré súvisia s tým, že všetky hodnoty, ktoré spadajú do sledovanej oblasti sú si rovnocenné, bez ohľadu na ich polohu v rámci sledovanej oblasti.



Jadrové analýzy hustoty – zohľadnenie váhy

- výpočet hustoty je možné rozšíriť o ďalší parameter, ktorý súvisí s tzv. geografickou váhou, alebo váhou vzdialenosti, ktorá sa použije ako koeficient na výpočet hodnôt.
- Použitie koeficientu na vyjadrenie váhy hodnoty na základe vzdialenosti je založený na rovnakom koncepte ako priestorová autokorelácia.
- dôležité je stanoviť si, akým spôsobom vyjadriť váhu hodnôt na základe vzdialenosti.
- vo všeobecnosti pre určenie hodnoty platí nasledujúci vzťah:

$$h_{(x,y)} = \sum_{i=1}^n \lambda_i p_i$$

λ_i – je váhový parameter.

Takýmto spôsobom je zabezpečené, aby hodnoty, ktoré sú bližšie k stredu sledovanej oblasti mali väčšiu váhu ako vzdialenejšie hodnoty.

všeobecný vzťah pre určenie váhy je možné zapísať vzťahom

$$\lambda_k(x) = \sum_{i=1}^n \frac{1}{\tau^2} k\left(\frac{x - x_i}{\tau}\right)$$

$\lambda_k(x)$ – váha, pre určenie hodnoty k v bode x , ktorý predstavuje polohu bodu (spravidla poloha v strede rastra), ku ktorej sa vzťahuje výpočet a ktorý tvorí zároveň stred sledovanej oblasti.

τ – je polomer vyhľadávania (bandwidth) sledovanej oblasti

$(x - x_i)$ – je vzdialenosť bodu x_i od stredu sledovanej oblasti x .

Jadrové analýzy hustoty – zohľadnenie váhy

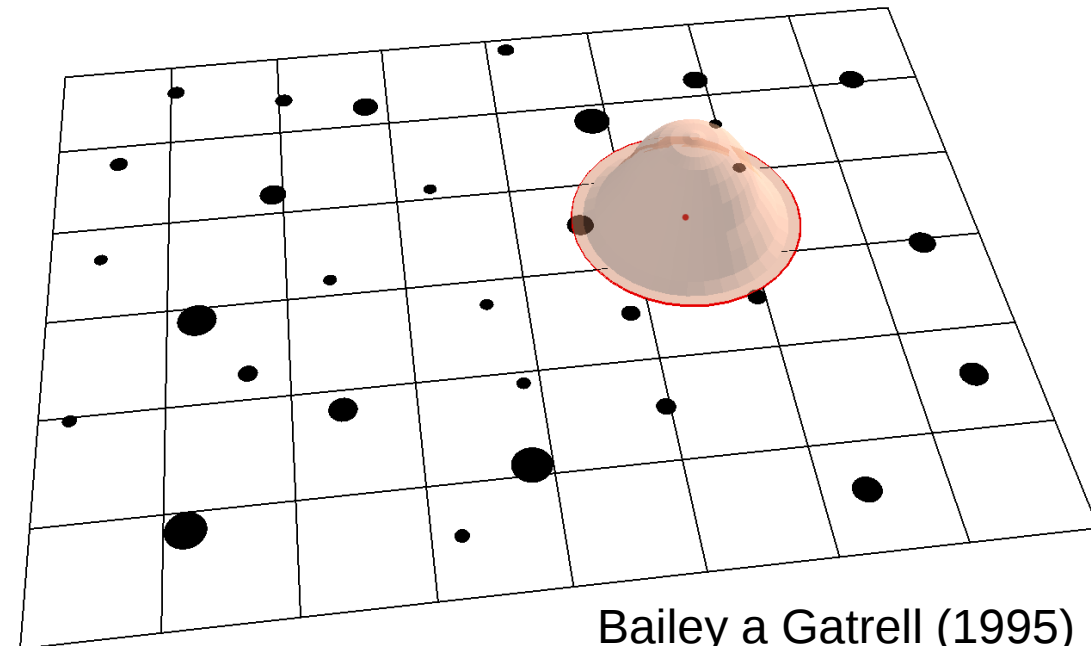
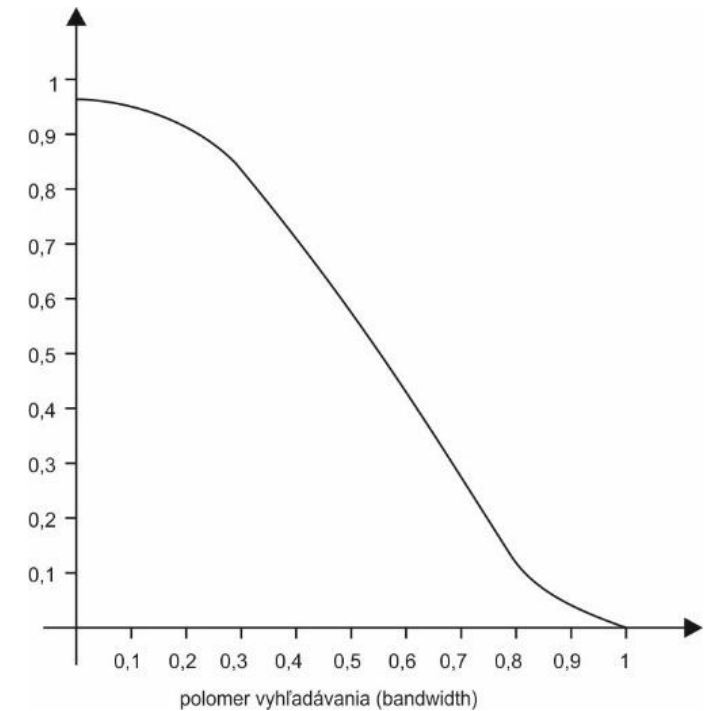
Existuje široká paleta rozličných prístupov ku tvorbe tzv. kernelu pre váženie.

Najčastejšie sa používa bikvadratická funkcia (quartic function, resp. polynomická funkcia 4. rádu)

Pre sledovanú oblasť (kernel) s tvarom kruhu platí nasledujúci vzťah:

$$\lambda_k(x) = \sum_{d_i \leq \tau} \frac{3}{\pi \tau^2} \left(1 - \frac{d_i^2}{\tau^2}\right)^2$$

d_i – je $(x - x_i)$, je vzdialenosť bodu x_i od stredu sledovanej oblasti (kernelu) x



Jadrové analýzy hustoty – zohľadnenie váhy

		x	y	$d_{(x,x_1)} = \sqrt{(x_x - x_{x_i})^2 + (x_y - x_{y_i})^2}$
Stred vyhľadávania	x	5,5	3,5	0
Bod s hodnotou 184	x ₁	4,569068	3,527496	0,799056
Bod s hodnotou 350	x ₂	6,093071	4,035497	0,931338

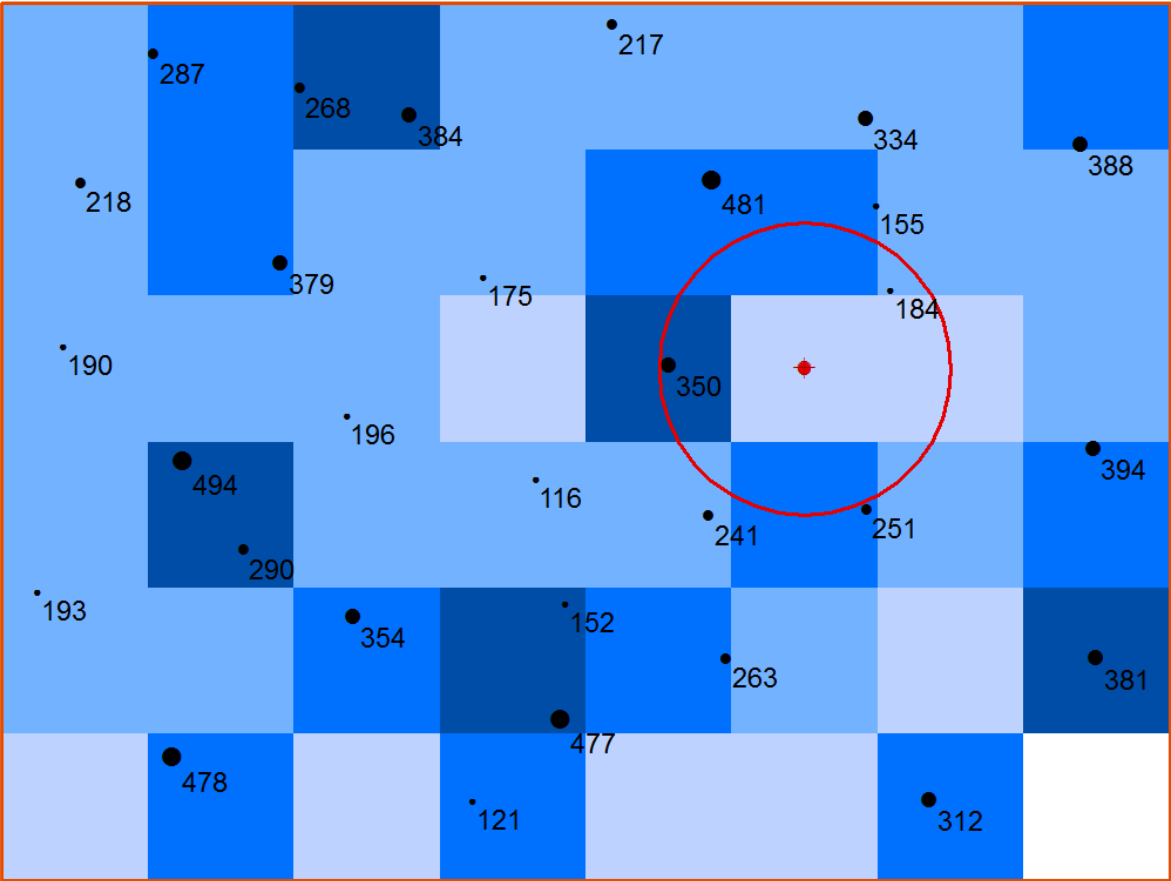
$$\lambda_k(x) = \sum_{d_i \leq \tau} \frac{3}{\pi \tau^2} \left(1 - \frac{d_i^2}{\tau^2}\right)^2$$

		$\frac{3}{\pi \tau^2}$	$\left(1 - \frac{d_i^2}{\tau^2}\right)^2$	$\lambda_k(x) = \frac{3}{\pi \tau^2} \left(1 - \frac{d_i^2}{\tau^2}\right)^2$
Bod s hodnotou 184	x ₁	0,954937	0,130689	0,1248
Bod s hodnotou 350	x ₂	0,954937	0,017585	0,016793

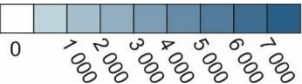
$$h_{(x,y)} = \sum_{i=1}^n \lambda_i p_i$$

		λ_i	p_i	$\sum_{i=1}^n \lambda_i p_i$
Bod s hodnotou 184	x ₁	0,954937	184	22,9632
Bod s hodnotou 350	x ₂	0,954937	350	5,877496
Σ				28,84069

```
ncols      8
nrows      6
xllcorner  0
yllcorner  0
cellsize   1
NODATA_value -9999
173,32 286,32 420,64 103,60 186,17 180,80 107,62 222,31
189,33 206,12 176,95 111,67 281,06 266,57 153,25 180,55
173,15 174,48 149,40 73,00 330,54 28,84 52,70 186,28
122,13 533,72 156,40 128,25 192,08 233,77 108,42 236,48
101,45 138,69 271,54 315,74 253,83 128,22 2,10 363,48
88,62 273,20 25,05 227,27 17,76 20,38 284,31 0
```



Jadrové analýzy hustoty - porovnanie

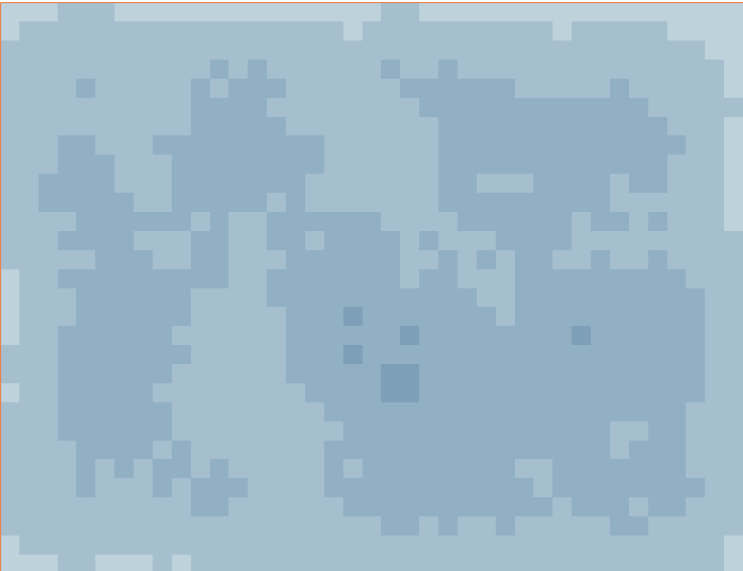
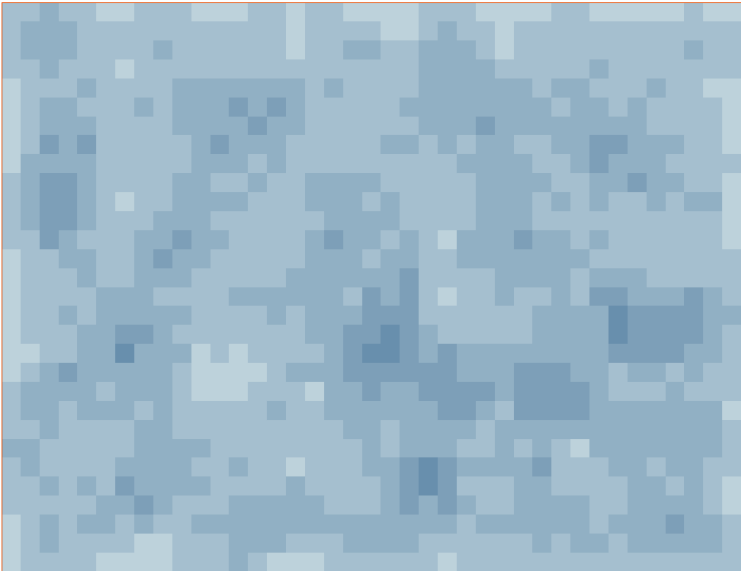
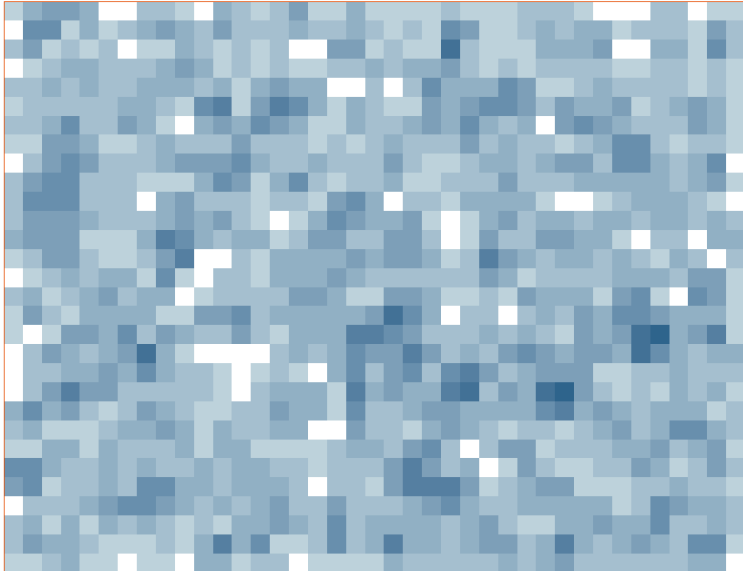


50 m

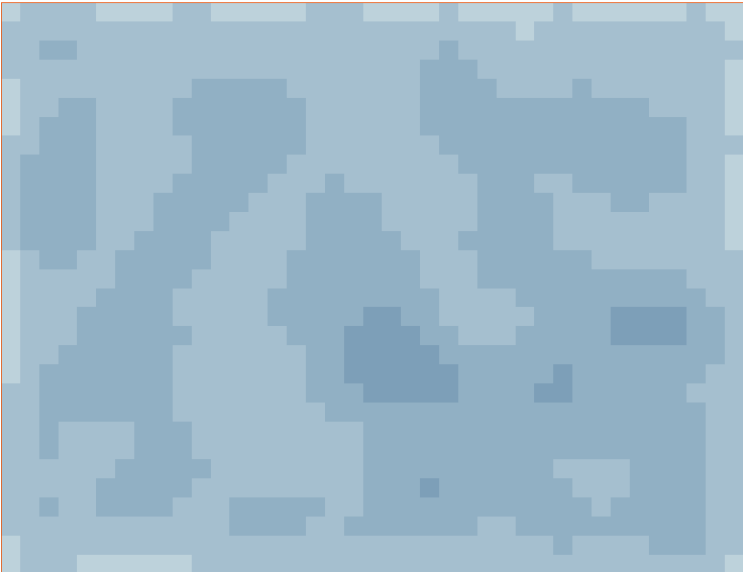
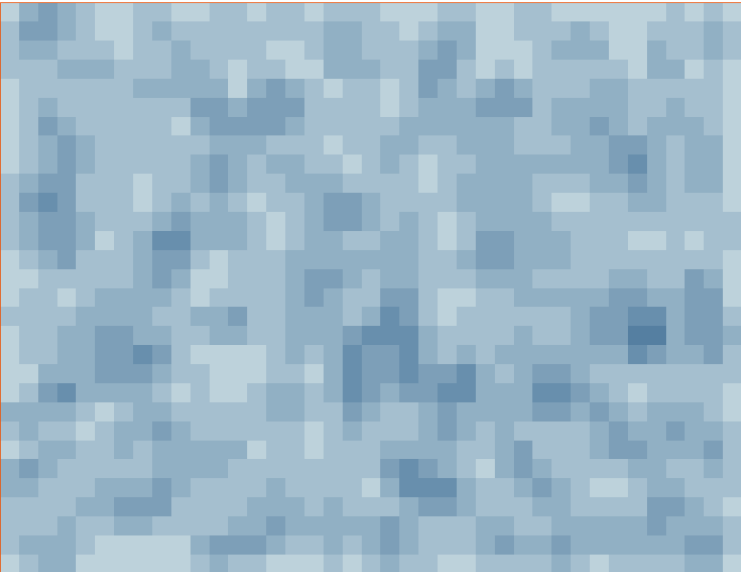
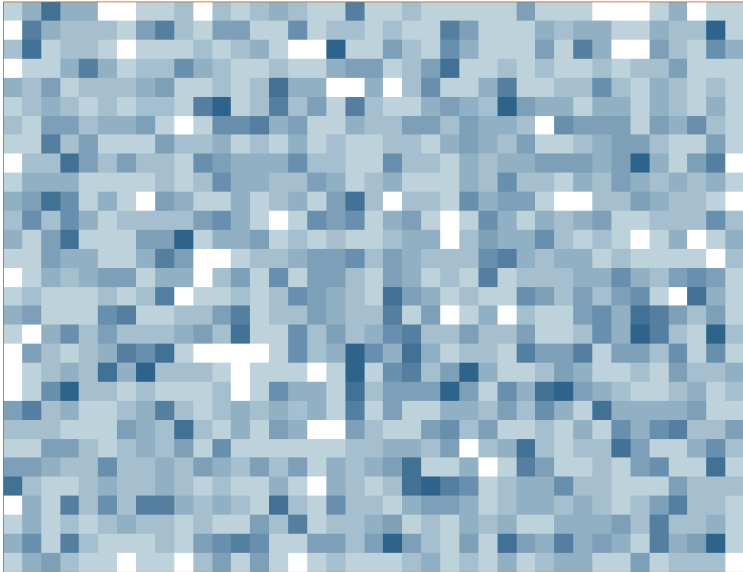
100 m

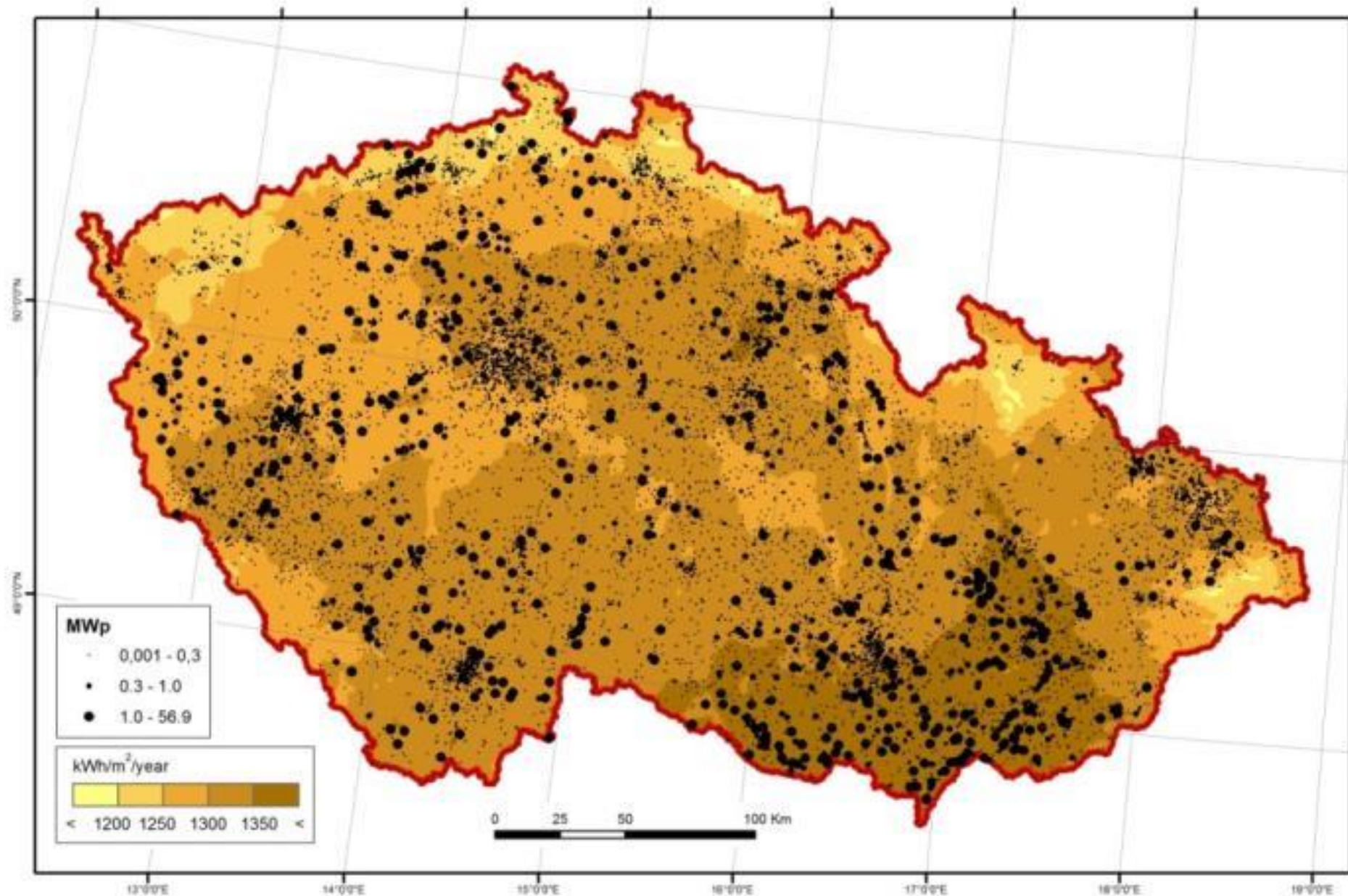
200 m

„population
parametrom“



„population
parameter“
váženie
vzdialenosťou



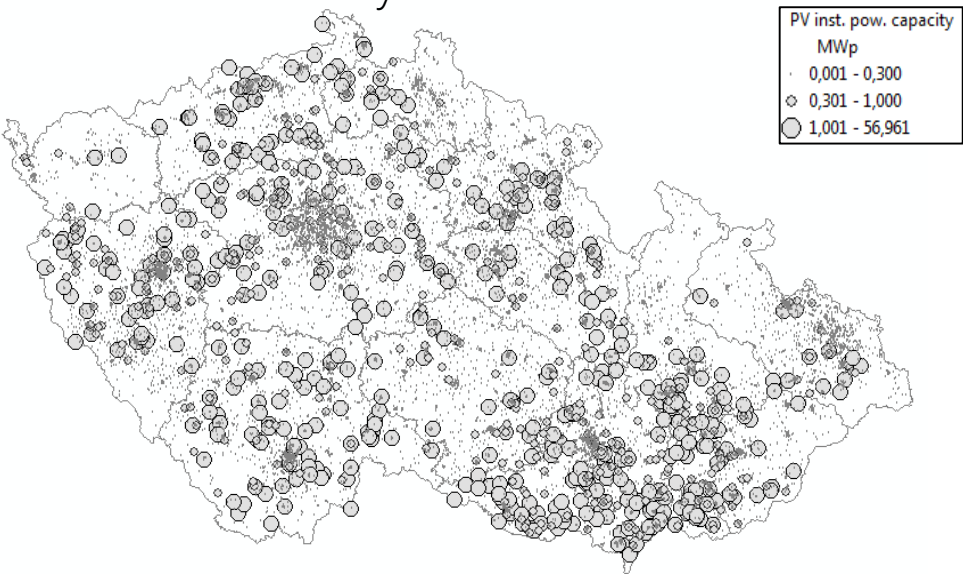


Prípadová štúdia – Kernel density estimation

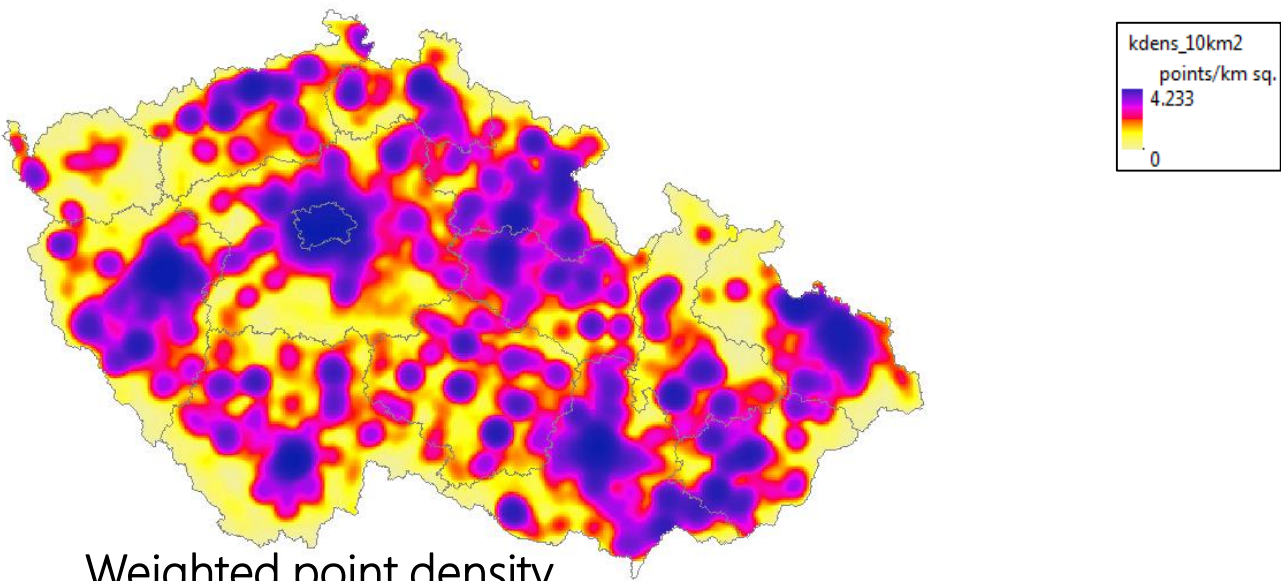
Poloha inštalácií PV elektrární



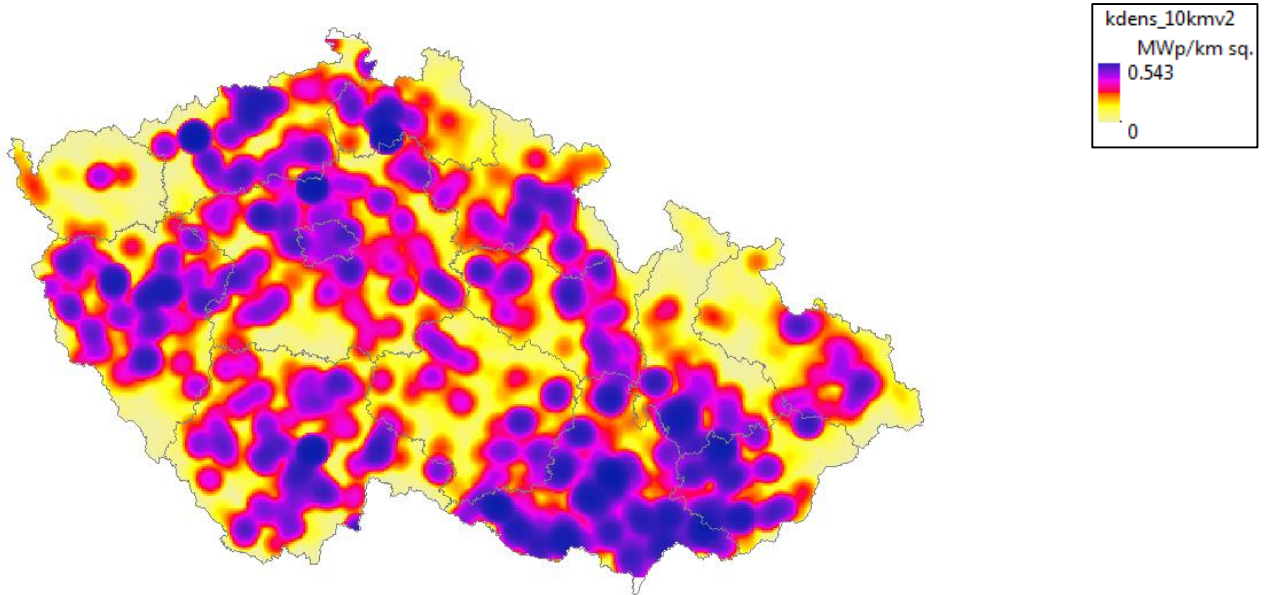
Graduated symbols



Hustota inštalácií

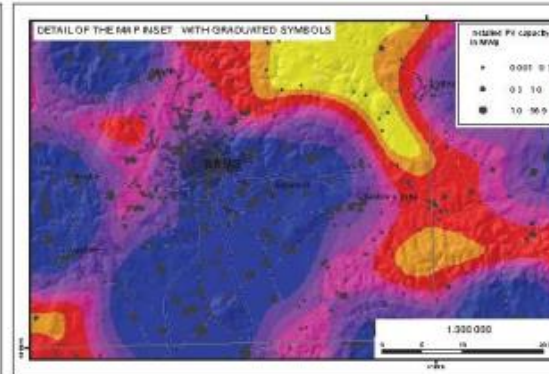
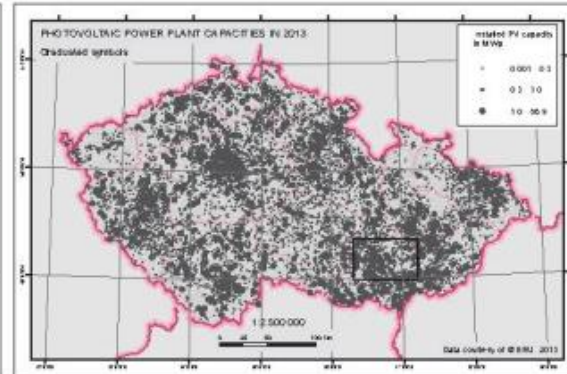
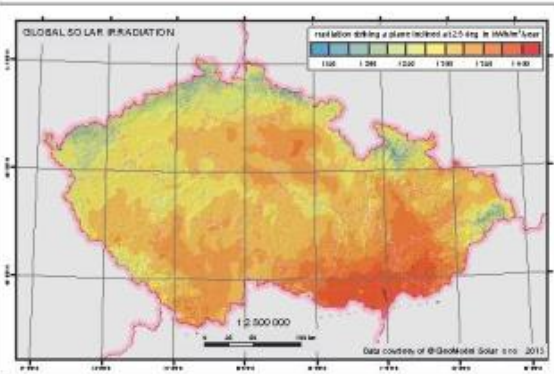
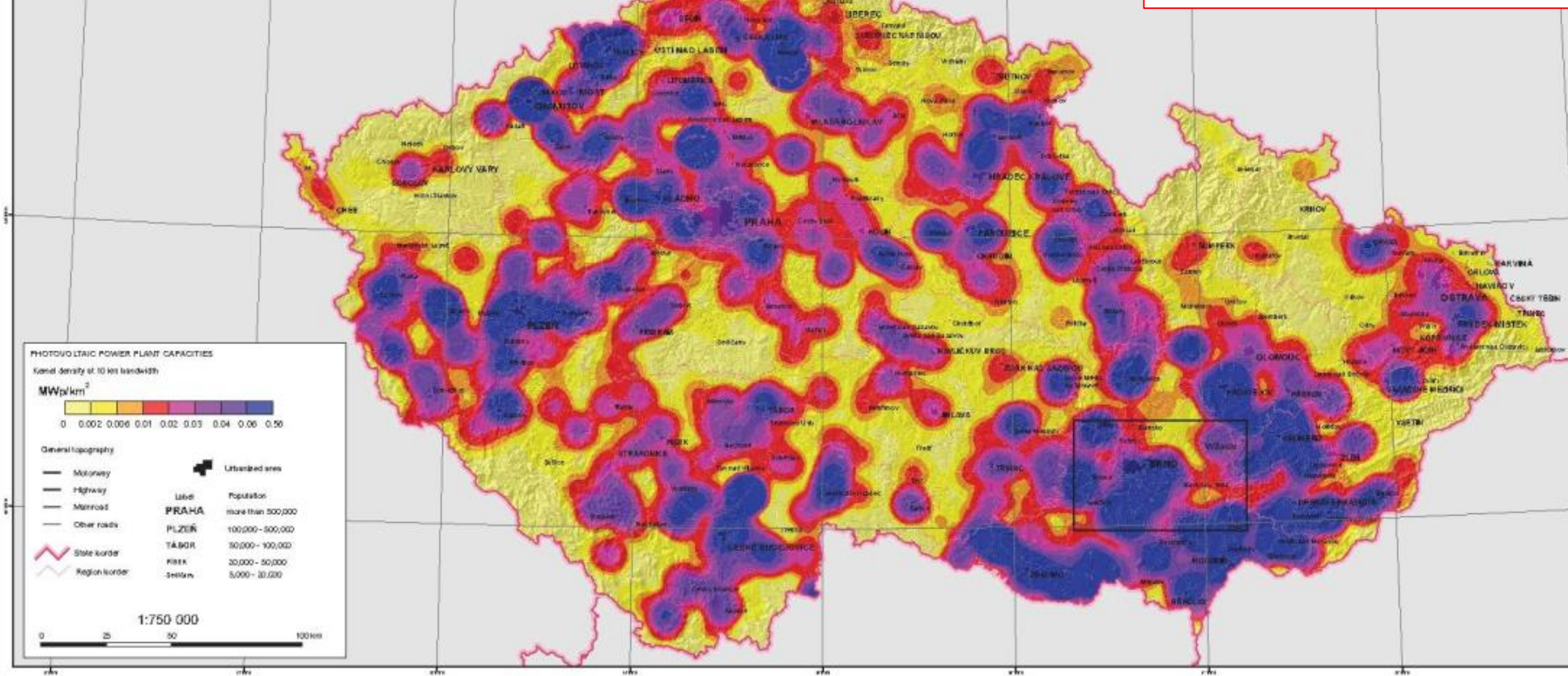


Weighted point density



CAPACITY OF PHOTOVOLTAIC POWER PLANTS IN THE CZECH REPUBLIC IN 2013

Gallay, M., Kaňuk, J., Hofierka, J. (2015). Capacity of photovoltaic power plants in the Czech Republic. *Journal of Maps*, 11(3), 480-486.



COORDINATE SYSTEM

WGS 1984 UTM Zone 32N
WKID: 32632 Authority: EPSG

Projection: Transverse Mercator
False Easting: 500000.0
False Northing: 0.0
Central Meridian: 15.0
Scale Factor: 0.9996
Latitude of Origin: 0.0
Units: Meter (1.0)

Geographic Coordinate System: GCS WGS 1984
Angular Unit: Degree (0.0174532925199433)
Prime Meridian: Greenwich (0.0)
Datum: G WGS 1984
Spheroid: WGS 1984
Semi-major Axis: 6378137.0
Semi-minor Axis: 6356752.314240179
Inverse Flattening: 298.257222501

General topography

Data source: ArcGIS via 3.1 GISDATA PRAHA

Designed for printing at A1

© Michal Gallay, Jan Kaňuk, Jaroslav Hofierka 2015

Michal Gallay, Jan Kaňuk, Jaroslav Hofierka
Institute of Geography
Faculty of Science
Pavol Jozef Šafárik University in Košice

© Journal of Maps 2015

Geograficky vážená regresia

Analýza lineárneho vzťahu (globálna)

- **Lineárna regresia** - analytické vyjadrenie lineárnej závislosti medzi dvomi alebo viacerými premennými

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

- určenie rovnice metódou najmenších štvorcov (ordinary least squares, OLS)

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} = \left(\frac{1}{n} \sum \mathbf{x}_i \mathbf{x}_i^T \right)^{-1} \left(\frac{1}{n} \sum \mathbf{x}_i y_i \right) \quad \mathbb{E}[\mathbf{x}_i \varepsilon_i] = 0$$

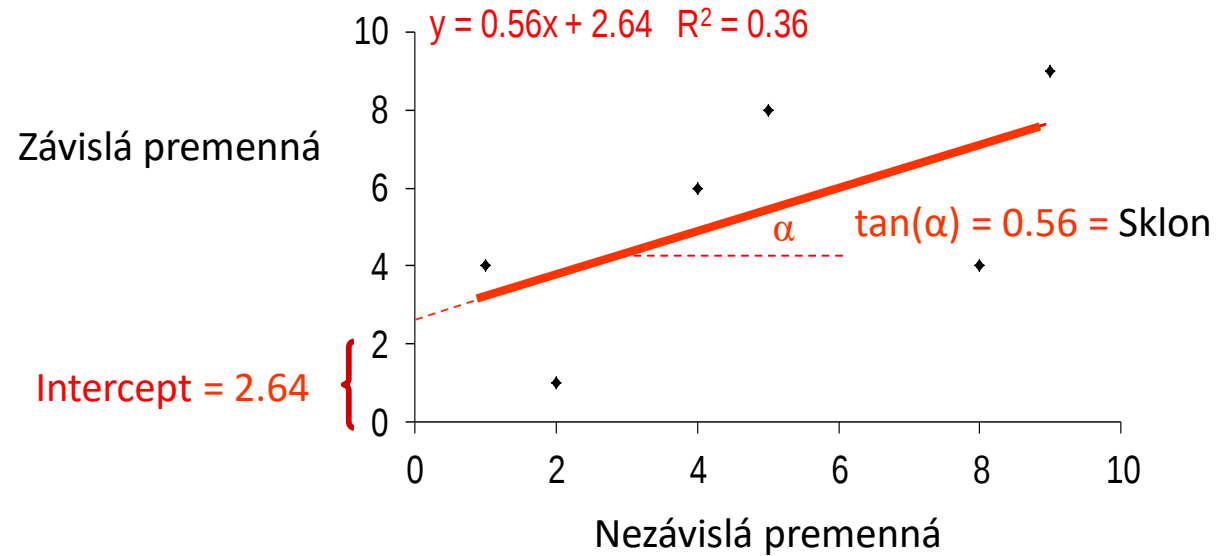
- všetky hodnoty v súbore vstupujú do výpočtu a majú rovnakú váhu.

Geograficky vážená regresia

Analýza lineárneho vzťahu (globálna)

Elevation (m)	Precipitation (mm)
46	12.10
98	29.40
50	18.20
32	14.80
24	12.70
113	13.10
123	18.50
25	10.30
16	28.80
260	27.00
100	16.75
283	45.40
250	25.70
262	47.50
18	22.65
248	24.70
254	50.50

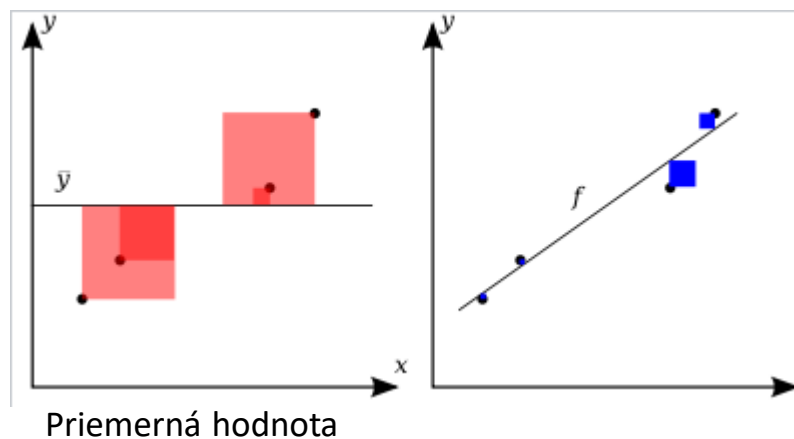
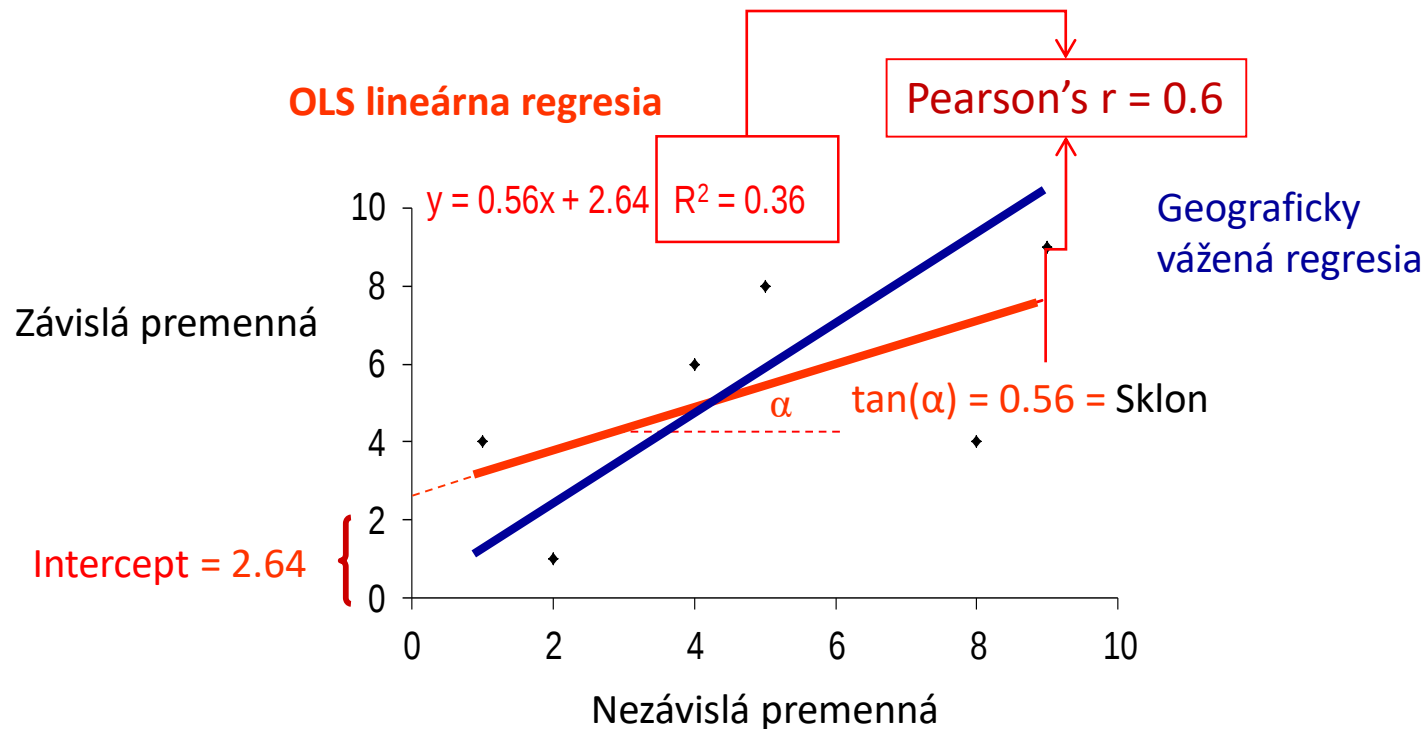
OLS lineárna regresia



Geograficky vážená regresia

Analýza lineárneho vzťahu (globálna)

Elevation (m)	Precipitation (mm)
46	12.10
98	29.40
50	18.20
32	14.80
24	12.70
113	13.10
123	18.50
25	10.30
16	28.80
260	27.00
100	16.75
283	45.40
250	25.70
262	47.50
18	22.65
248	24.70
254	50.50



Geograficky vážená regresia

Geographically Weighted Regression

Analýza lineárneho vzťahu (lokálna)

X (m)	Y (m)	Elevation (m)	Precipitation (mm)
426100	420000	46	12.10
429300	423300	98	29.40
422000	422500	50	18.20
432000	419600	32	14.80
434600	420400	24	12.70
427300	408400	113	13.10
424600	432400	123	18.50
432300	431500	25	10.30
439600	423900	16	28.80
421300	405600	260	27.00
427200	437500	100	16.75
415300	405800	283	45.40
409600	413000	250	25.70
414100	405800	262	47.50
444400	426500	18	22.65
407200	425600	248	24.70
410900	406800	254	50.50

Chris Lloyd (2007): Local models for Spatial Analysis. CRC Press. 244 p.

Geograficky vážená regresia

Geographically Weighted Regression

Analýza lineárneho vzťahu (lokálna)

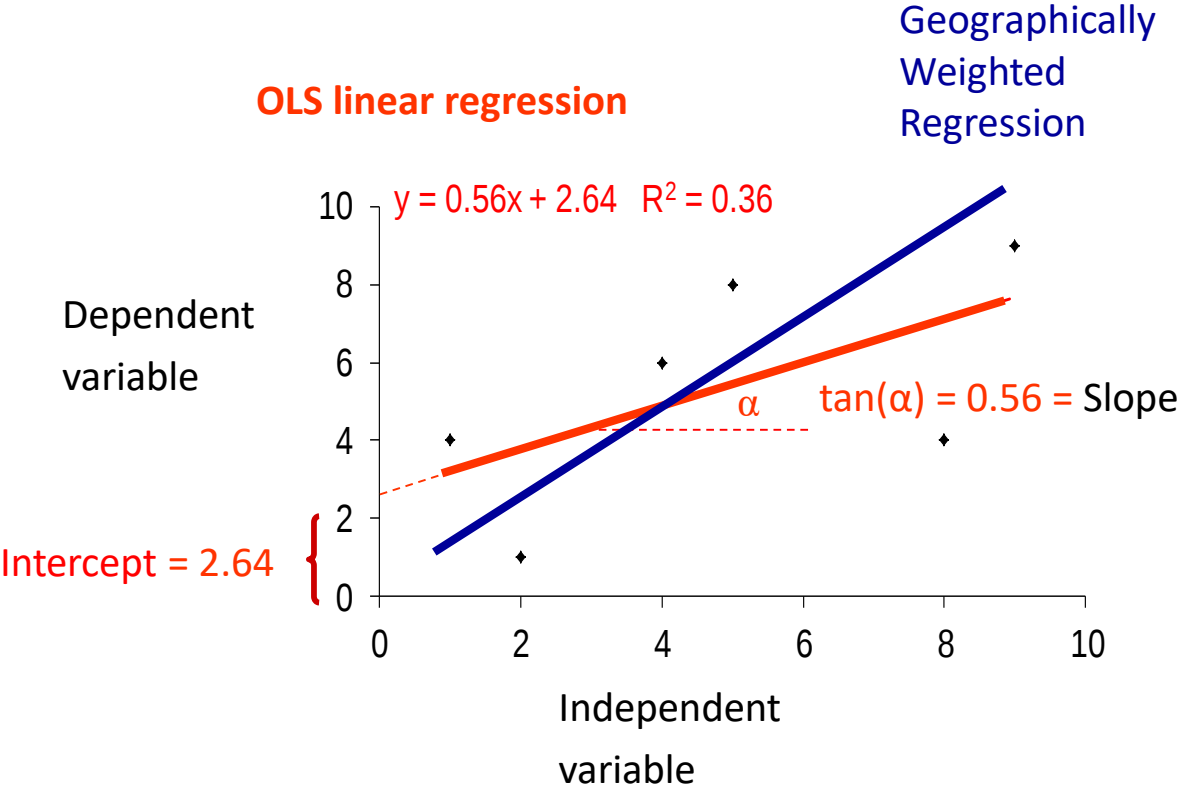
X (m)	Y (m)	Elevation (m)	Precipitation (mm)	Distance (m)	Weight
426100	420000	46	12.10	0	1.000
429300	423300	98	29.40	4596.738	0.917
422000	422500	50	18.20	4802.083	0.910
432000	419600	32	14.80	5913.544	0.867
434600	420400	24	12.70	8509.407	0.744
427300	408400	113	13.10	11661.904	0.574
424600	432400	123	18.50	12490.396	0.529
432300	431500	25	10.30	13064.838	0.499
439600	423900	16	28.80	14052.046	0.447
421300	405600	260	27.00	15178.933	0.391
427200	437500	100	16.75	17534.537	0.285
415300	405800	283	45.40	17840.404	0.273
409600	413000	250	25.70	17923.448	0.270
414100	405800	262	47.50	18591.396	0.244
444400	426500	18	22.65	19420.093	0.215
407200	425600	248	24.70	19712.179	0.205
410900	406800	254	50.50	20131.567	0.192

Chris Lloyd (2007): Local models for Spatial Analysis. CRC Press. 244 p.

Geograficky vážená regresia

Analýza lineárneho vzťahu (lokálna)

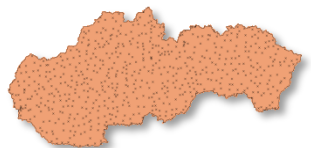
X (m)	Y (m)	Elevation (m)	Precipitation (mm)	Distance (m)	Weight
426100	420000	46	12.10	0	1.000
429300	423300	98	29.40	4596.738	0.917
422000	422500	50	18.20	4802.083	0.910
432000	419600	32	14.80	5913.544	0.867
434600	420400	24	12.70	8509.407	0.744
427300	408400	113	13.10	11661.904	0.574
424600	432400	123	18.50	12490.396	0.529
432300	431500	25	10.30	13064.838	0.499
439600	423900	16	28.80	14052.046	0.447
421300	405600	260	27.00	15178.933	0.391
427200	437500	100	16.75	17534.537	0.285
415300	405800	283	45.40	17840.404	0.273
409600	413000	250	25.70	17923.448	0.270
414100	405800	262	47.50	18591.396	0.244
444400	426500	18	22.65	19420.093	0.215
407200	425600	248	24.70	19712.179	0.205
410900	406800	254	50.50	20131.567	0.192



Geograficky vážená regresia

Globálny model

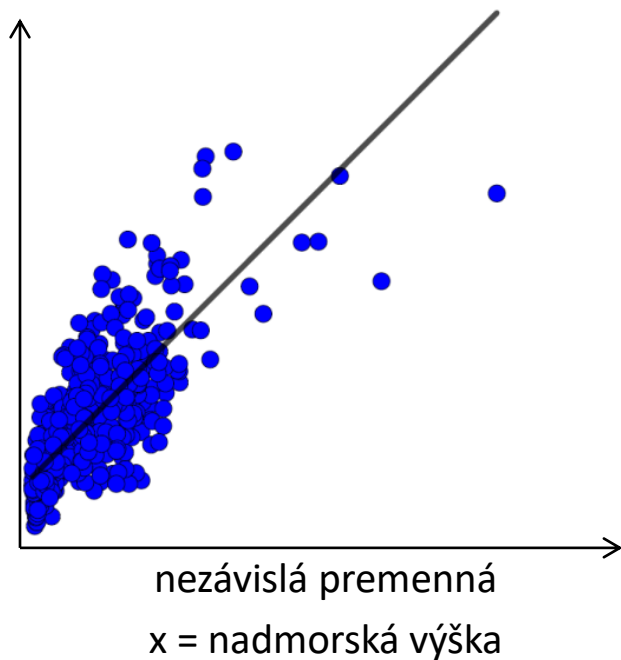
■ Lineárna regresia (OLS)



$$y = 0,42 \cdot x + 545,24; r^2 = 0,59; r = 0,77$$

SD = 62,35 mm

závislá premenná
 y = úhrn zrážok



$$y = X\beta + \varepsilon,$$

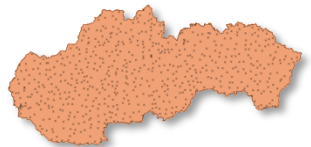
ZRAZKY	VYSKA
513	124
519	116
523	110
531	121
535	138
537	123
538	169
539	152
541	125
541	148

y = úhrn zrážok
závislá premenná
 x = nadmorská výška
nezávislá premenná

Geograficky vážená regresia

Globálny model

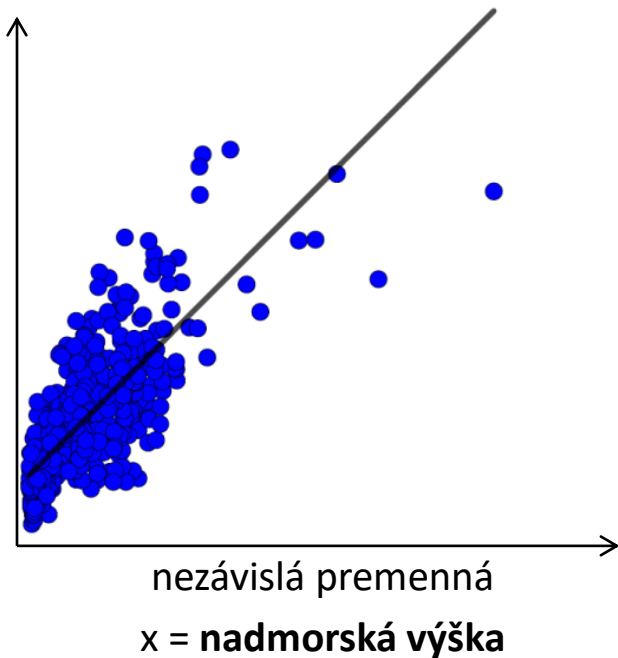
- Lineárna regresia (OLS)



$$y = 0,42 \cdot x + 545,24; r^2 = 0,59; r = 0,77$$

SDE = 62,35 mm

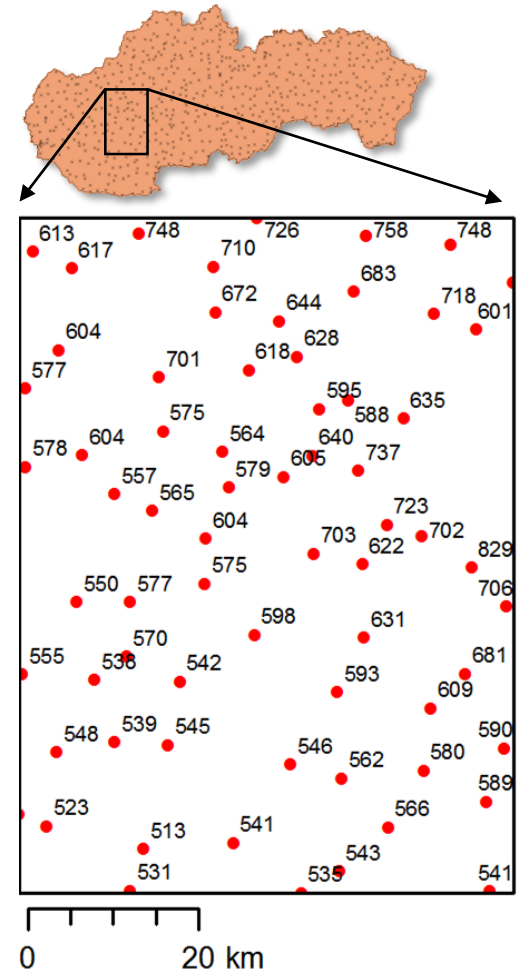
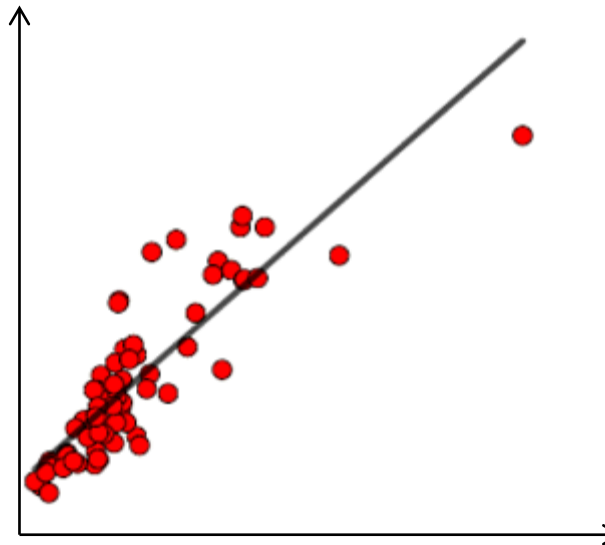
závislá premenná
 y = úhrn zrážok



Lokálny model

- Lokálna regresia na podmnožine

$$y = 0,791 \cdot x + 446,81 \quad r^2 = 0,76 \quad r = 0,88 \quad \text{SDE} = 35,42 \text{ mm}$$



Geograficky vážená regresia

Lokálny model

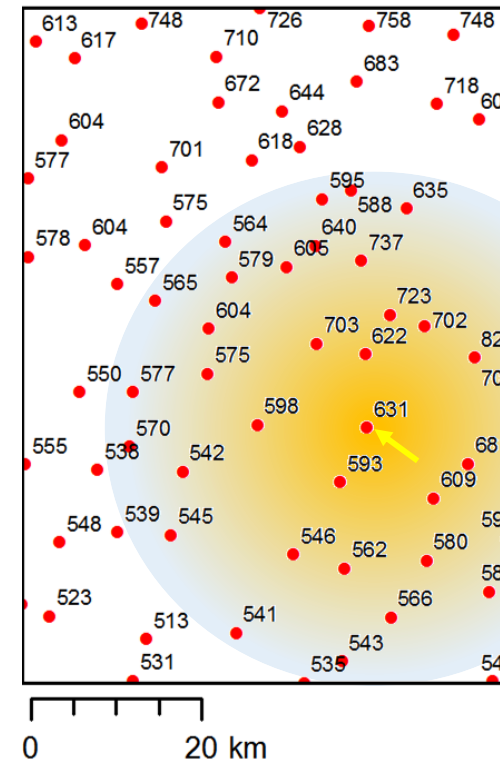
- geograficky vážená regresia (angl. GWR), Fotheringham et al. (2000), Lloyd (2009)

SHMU	ZRAZKY	VYSKA	X-JTSK	Y-JTSK	DISTANCE	G_WEIGHTH
ZLATE_MORAVCE	631	203	-475551	-1263423	0	1,0000
ZIKAVA	622	294	-475760	-1254800	8625	0,9482
SKYCOV	723	409	-472890	-1250165	9182	0,9415
KRNCA	605	220	-484990	-1244518	9726	0,9346
ZLATNO	703	329	-481419	-1253565	11472	0,9101
KLATOVA_NOVA_VES	640	198	-481629	-1242048	11519	0,9094
JELENEC	598	191	-488449	-1263097	11843	0,9045
NITRIANSKA_STREDA	579	173	-491438	-1245757	12702	0,8910
HOSTIE,MALE_DUBOVO	702	316	-468761	-1251416	12840	0,8887
OPONICE	604	168	-494146	-1251751	12855	0,8885
HORNE_LEFANTOVCE	575	190	-494239	-1257044	13283	0,8814
VELKA_LEHOTA	829	589	-462923	-1255175	15082	0,8498
TOPOLCANY	564	174	-492186	-1241599	16096	0,8308
TESARSKE_MLYNANY	593	196	-478714	-1269819	16478	0,8234
VRABLE	546	142	-484217	-1278339	17252	0,8082
PARTIZANSKE-MALE_BIELICE	588	195	-477452	-1235492	18503	0,7827
HORNE_OBDOKOVCE	565	178	-500370	-1248474	19622	0,7592
NITRA	542	173	-497212	-1268564	21780	0,7122

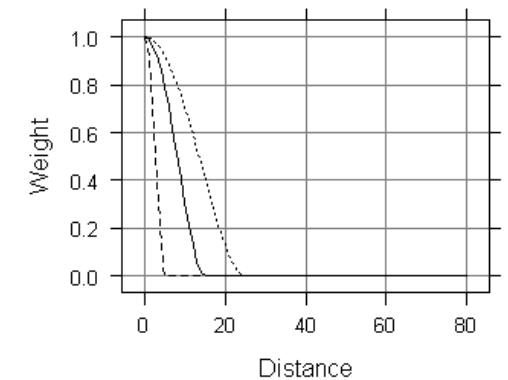
Bandwidth τ / Cutoff τ

- 5 m
- 15 m
- 25 m

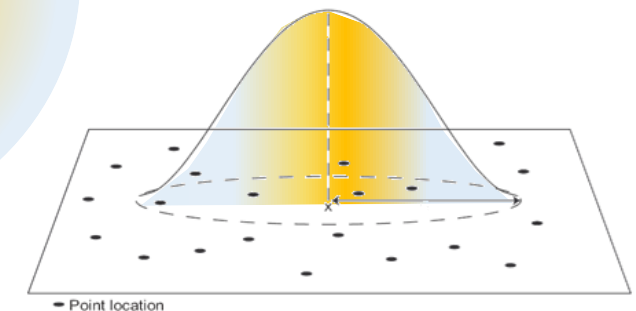
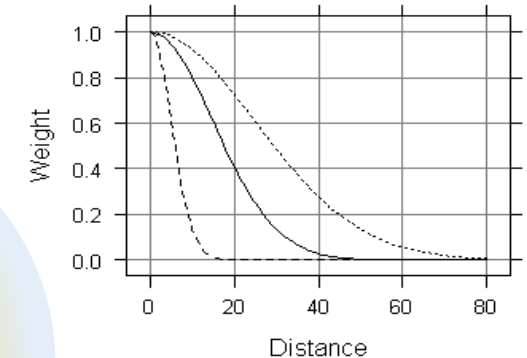
$$w_i(x) = \exp[-0.5(d/\tau)^2]$$



Bisquare function with a cutoff



Gaussova váhová funkcia

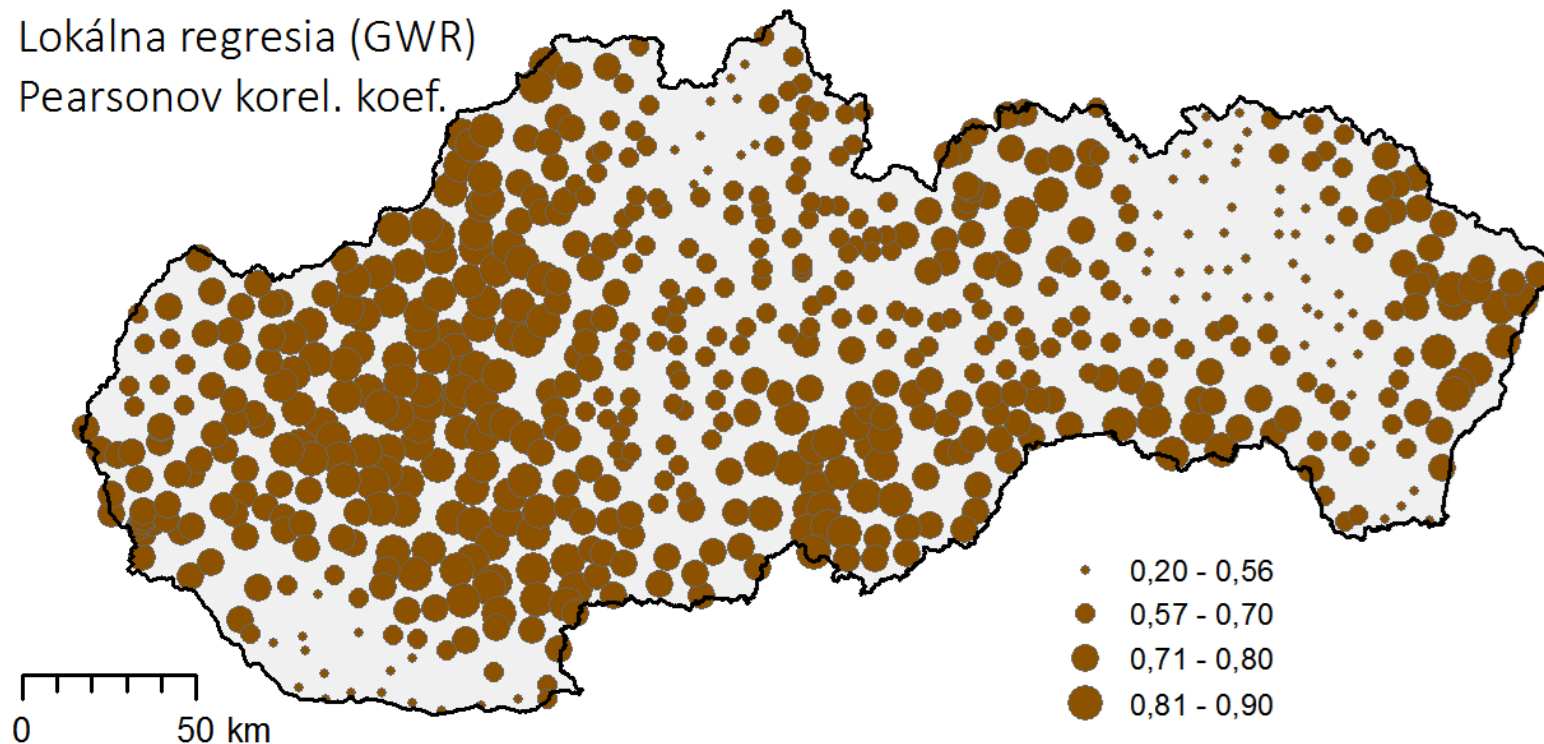


Geograficky vážená regresia

Lokálny model

■ geograficky vážená regresia

Lokálna regresia (GWR)
Pearsonov korel. koef.



Parametre globálnej lineárnej regresie
(OLS)

$$y = 545,24 + 0,42 \cdot x ; r^2 = 0,59; \mathbf{r = 0,76}$$

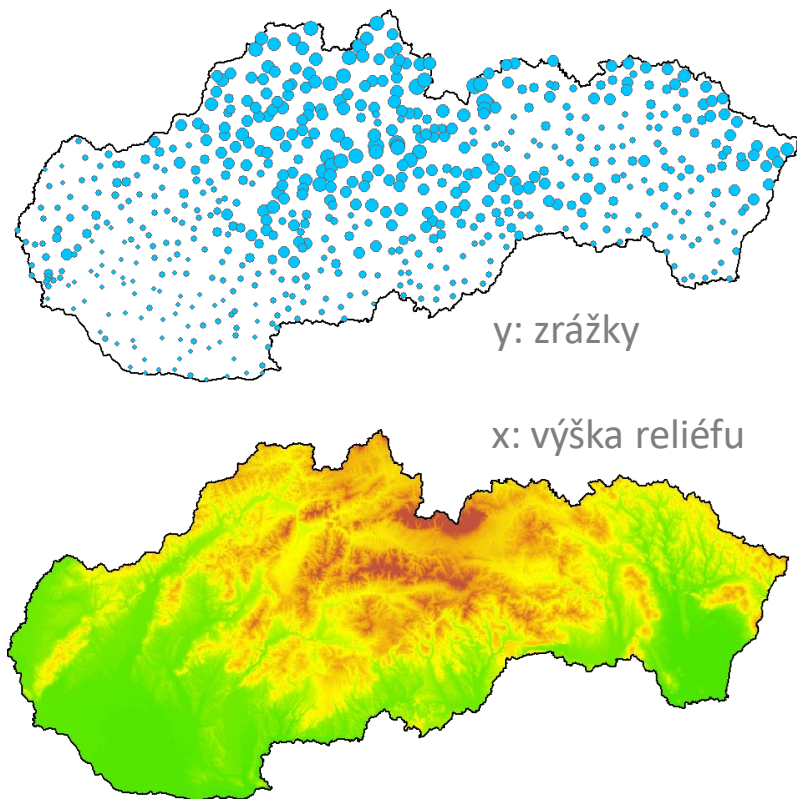
SDE = 62,35 mm

x = nadmorská výška

y = úhrn zrážok

Geograficky vážená regresia

Predikcia s použitím globálneho modelu vzťahu úhrnu zrážok a nadmorskej výšky reliéfu



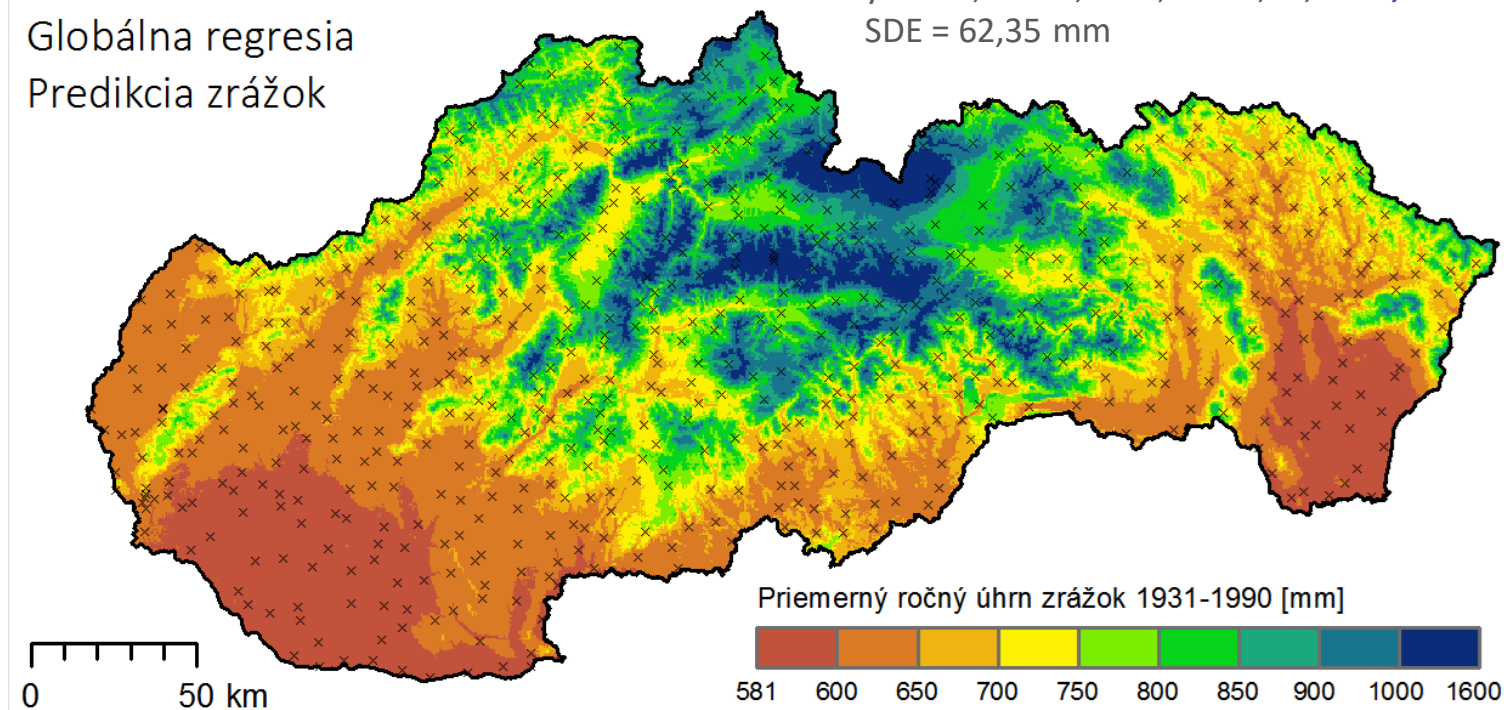
Globálna regresia
Predikcia zrážok

0 50 km

Parametre globálnej lineárnej regresie
(OLS)

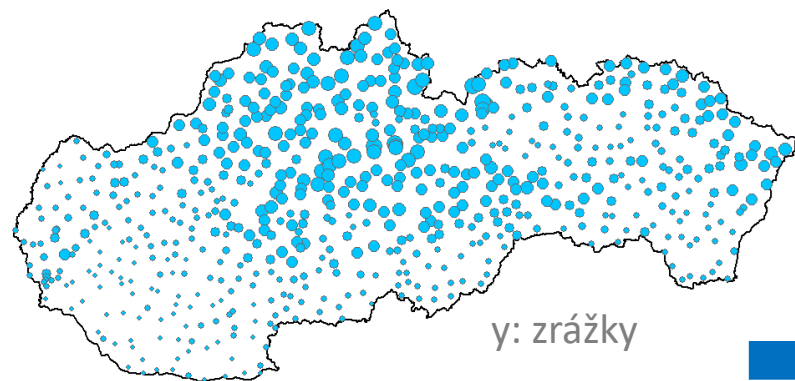
$$y = 545,24 + 0,42 \cdot x ; r^2 = 0,59; \mathbf{r = 0,76}$$

SDE = 62,35 mm



Geograficky vážená regresia

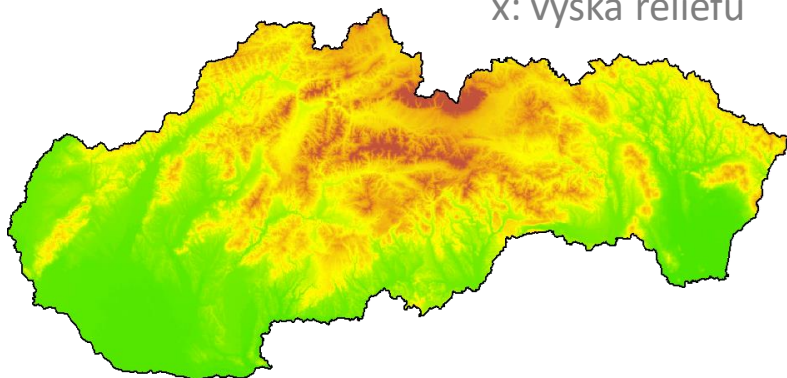
Predikcia s použitím GWR modelu
vzťahu úhrnu zrážok a nadmorskej výšky reliéfu



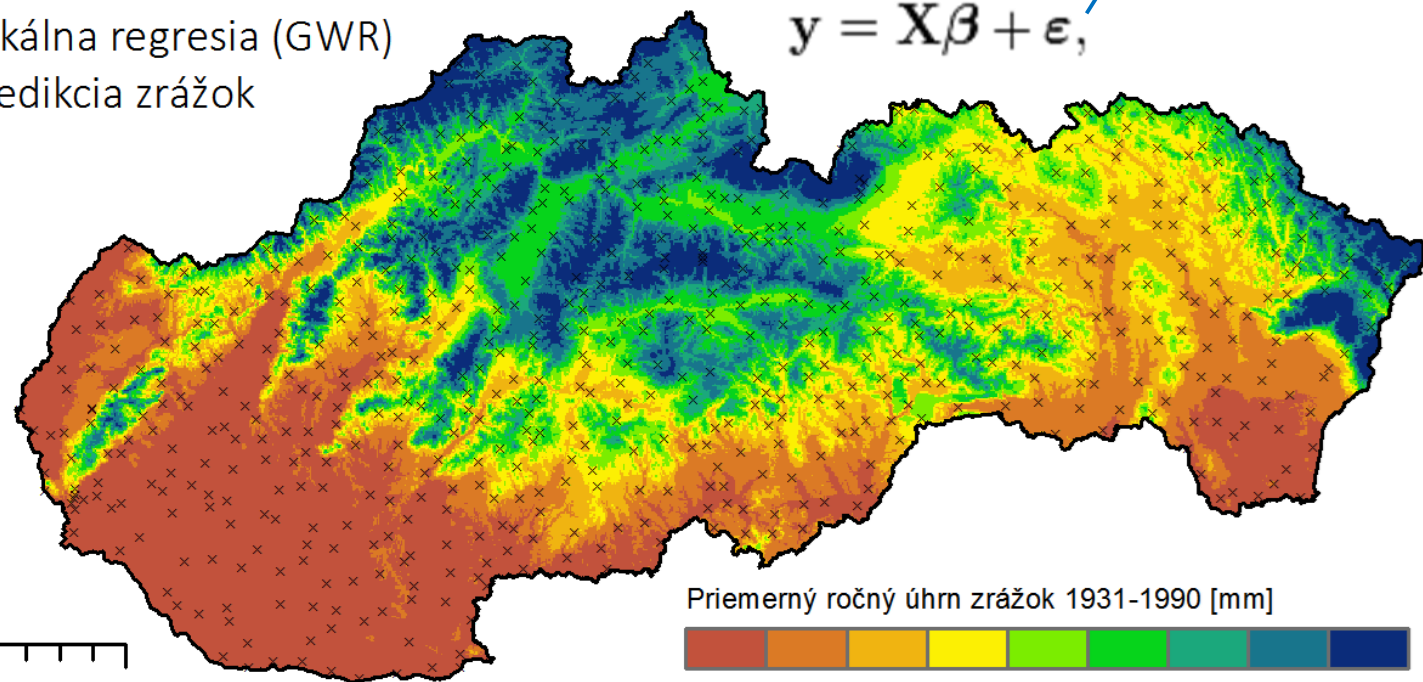
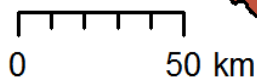
y: zrážky



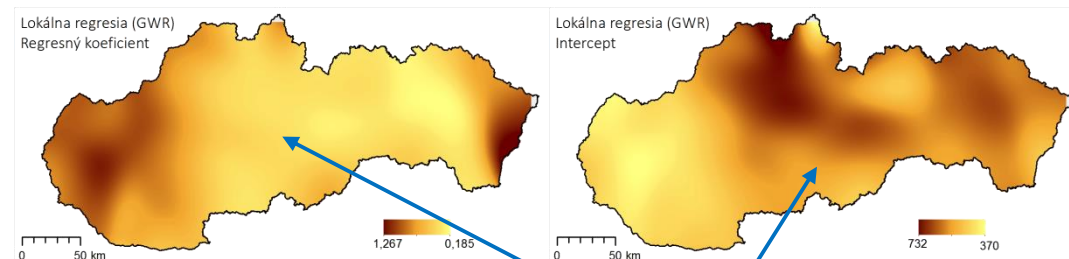
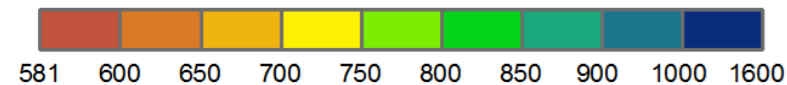
x: výška reliéfu



Lokálna regresia (GWR)
Predikcia zrážok



Priemerný ročný úhm zrážok 1931-1990 [mm]

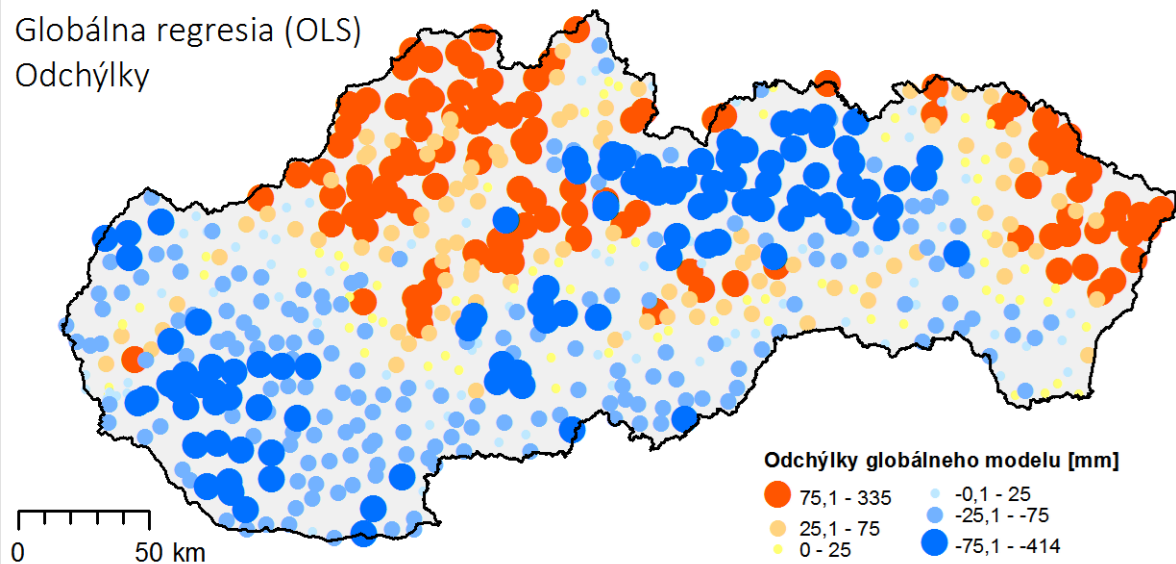


$$y = X\beta + \epsilon,$$

Geograficky vážená regresia

Predikcia : porovnanie odchýlok globálneho a lokálneho regresného modelu
lineárny vzťah úhrnu zrážok a nadmorskej výšky reliéfu

Globálna regresia (OLS)
Odchýlky



Lokálna regresia (GWR)
Odchýlky

